Estimating the Quadratic Covariation from Asynchronous Noisy High-Frequency Observations

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

Dr. Rer. Nat.
im Fach Mathematik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät II
Humboldt-Universität zu Berlin

von

Dipl.-Math. Markus Bibinger
geboren am 24.08.1981 in Ulm

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät II:
Prof. Dr. Elmar Kulke

Gutachter:
1. Prof. Dr. Markus Reiß
2. Prof. Dr. Vladimir Spokoiny
3. Prof. Dr. Yacine Aït-Sahalia

eingereicht am: 30.03.2011
Tag der mündlichen Prüfung: 15.07.2011

Abstract

A nonparametric estimation approach for high-frequency observations from Itô processes with an additive noise is developed. We consider a bivariate model with asynchronous observation schemes and correlated Brownian motions. The goal is to find a good estimator for the quadratic covariation of the two Itô processes that paves the way for statistical inference.

It is proved that a closely related sequence of statistical experiments is locally asymptotically normal (LAN) in the Le Cam sense. By virtue of this property optimal convergence rates and efficiency bounds for asymptotic variances of estimators can be concluded.

The proposed nonparametric estimator is founded on a combination of two modern estimation methods devoted to an additive observation noise on the one hand and asynchronous observation schemes on the other hand. Both are motivated and introduced at first to provide the grounding on that the combined estimator for the general model can be constructed. With the inflow of the theory by Hayashi and Yoshida on the estimation problem for non-synchronous observations and influences from other authors, we reinvent this Hayashi-Yoshida estimator in a new illustration that can serve as a synchronization method which is possible to adapt for the combined approach. A stable central limit theorem is proved focusing especially on the impact of characteristics of non-synchronicity on the asymptotic variance.

With this preparations on hand, the generalized multiscale estimator for the noisy and asynchronous setting arises. This convenient method for the general model is based on subsampling and multiscale estimation techniques that have been established by Mykland, Zhang and Aït-Sahalia. It preserves valuable features of the synchronization methodology and the estimators to cope with noise perturbation. The central result of the thesis is that the estimation error of the generalized multiscale estimator converges with optimal rate stably in law to a centred mixed normal limiting distribution on fairly general regularity assumptions.

For the asymptotic variance a consistent estimator based on time transformed histograms is given making the central limit theorem feasible. In an application study a practicable estimation algorithm including a choice of tuning parameters is tested for its features and finite sample size behaviour. We take account of recent advances on the research field by other authors in comparisons and notes.

Zusammenfassung

Ein nichtparametrisches Schätzverfahren für hochfrequente Beobachtungen von Itô-Prozessen mit einem additiven Rauschen wird entwickelt. Zugrunde liegt ein bivariates statistisches Modell mit nicht-synchronen Beobachtungsschemata und korrelierten Brownschen Bewegungen. Das Ziel ist einen geeigneten Schätzer für die quadratische Kovariation der Itô-Prozesse herzuleiten, welcher zudem den Weg zur statistischen Inferenz ebnet.

Für eine artverwandte Folge von statistischen Experimenten wird die lokal asymptotische Normalität (LAN) im Sinne von Le Cam bewiesen. Mit dieser lassen sich optimale Konvergenzraten und Effizienzschranken für asymptotische Varianzen von Schätzern ableiten. Der in dieser Arbeit vorgestellte nichtparametrische Schätzer wird auf Grundlage von zwei modernen Schätzverfahren, für die Anwendung bei nicht-synchronen Beobachtungen zum einen, und einem additiven Rauschen zum anderen, entwickelt. Diese beiden werden zunächst motiviert und eingeführt um das Fundament zu schaffen, auf dem aufbauend dann das kombinierte Verfahren konstruiert werden kann. Mit Hilfe des Einflusses der Theorie von Hayashi und Yoshida zu dem Schätzproblem bei nicht-synchronen Beobachtungen und weiterer Einflüsse anderer Autoren, wird der Hayashi-Yoshida Schätzer in einer neuen Darstellung eingeführt, welche einen Synchronisierungsalgorithmus mit einschließt, der für die kombinierte Methode ausgelegt werden kann. Es wird eine stabiles zentrales Grenzwerttheorem bewiesen, wobei spezieller Wert auf die Analyse des Einflusses bestimmter Eigenschaften der Nicht-Synchronität auf die asymptotische Varianz gelegt wird.

Nach diesen Vorbereitungen kann mit den entsprechenden Methoden das kombinierte Schätzverfahren vorgestellt werden. Dieses für den allgemeinsten Fall nicht-synchroner verrauschter Beobachtungen passende Verfahren beruht auf Subsampling- und Multiskalenmethoden, die auf Mykland, Zhang und Aït-Sahalia zurück gehen. Es vereint positive Eigenschaften der beiden Ursprünge. Das zentrale Resultat dieser Arbeit ist der Beweis, dass der Schätzfehler des sogenannten verallgemeinerten Multiskalenschätzers stabil in Verteilung gegen eine zentrierte gemischte Normalverteilung konvergiert. Für die asymptotische Varianz wird ein konsistenter Schätzer unter Verwendung von zeittransformierten Histogrammen angegeben wodurch das stabile Konvergenztheorem nutzbar wird. In einer Anwendungsstudie wird eine praktische Implementierung des Schätzverfahrens, die die Wahl von abhängigen Parametern beinhaltet, getestet und auf ihre Eigenschaften im Falle endlicher Stichprobenumfänge untersucht. Neuen fortgeschrittenen Entwicklungen auf dem Forschungsfeld von Seite anderer Autoren wird Rechnung getragen durch Vergleiche und diesbezügliche Kommentare.

Contents

Introduction

A central model in statistics is the regression model of the type $Y = f(X) + \epsilon$. The function f describes a deterministic relation between the observable quantity Y and the signal X. The random noise ϵ is independent of X and captures the deviance from a purely deterministic relation. Scientists from various fields work with the regression model since the setting of noisy indirect measurements of a quantity of interest serves often as a useful approximation of the truth.
In addition to the statistical standard regression setup, the theory of the thesis on hand is founded on the notion of stochastic processes evolving in continuous time. These have become of great importance for modeling dynamics of systems in different research areas since Kiyoshi Itô originated the theory of stochastic analysis and stochastic differential equations in the 1940s.
Imagine a bivariate regression type model with a non-deterministic signal that follows continuous-time stochastic processes X and Y. For these processes indirect observations with a random perturbation are available:

$$\tilde{X}_{t_i} = X_{t_i} + \epsilon_{t_i}^X \ , i \in \{0, \dots, n\}\,, \tag{0.1a}$$

$$\tilde{Y}_{\tau_j} = Y_{\tau_j} + \epsilon_{\tau_j}^Y \ , j \in \{0, \dots, m\}\,. \tag{0.1b}$$

The combination with the problem that in such a multivariate model discrete observations may take place at asynchronous times sets up the framework of this thesis. It is a challenging issue to develop and analyze asymptotic statistical methods for the additive noise model and it will be material to incorporate some sophisticated statistical concepts. Since the problems of noise perturbation and non-synchronicity, each being interesting from a mathematical point of view on its own, often coincide in applications, it is a crucial task to provide statistical solutions to the joint problem.

Statistical model and the estimation problem

The goal of this work is to provide a nonparametric estimation approach for the quadratic covariation of the two Itô processes (0.1a) and (0.1b) at a fixed time $T < \infty$ in the bivariate statistical model of discrete observations on the time span $[0, T]$ where the two Itô processes are latent. This means they are observed with additive noise. The Itô processes X and Y, which we call efficient processes in this context, are solutions of the stochastic differential equations

$$dX_t = \int_0^t \mu_t^X \, dt + \int_0^t \sigma_t^X \, dB_t^X \ , \ dY_t = \int_0^t \mu_t^Y \, dt + \int_0^t \sigma_t^Y \, dB_t^Y \ ,$$

with two standard Brownian motions B^X and B^Y, locally bounded random processes μ^X and μ^Y, that we call drift processes, and random processes σ^X and σ^Y with continuous paths, that we call volatility processes. We presume that $d\langle B^X, B^Y\rangle_t = \rho_t\, dt$ for a continuous process ρ, where $\langle B^X, B^Y\rangle$ denotes the quadratic covariation of the Brownian motions. We focus on estimating the quadratic covariation of the processes X and Y: $\langle X, Y\rangle_T = \int_0^T \rho_t \sigma_t^X \sigma_t^Y\, dt$.

In the setting of $(n+1)$ synchronous observation times $t_i^{(n)}$, it is a well-known result that the stochastic convergences

$$\sum_{i=1}^n \left(X_{t_i^{(n)}} - X_{t_{i-1}^{(n)}}\right)\left(Y_{t_i^{(n)}} - Y_{t_{i-1}^{(n)}}\right) \xrightarrow{p} \langle X, Y\rangle_T \quad \text{and} \quad \sum_{i=1}^n \left(X_{t_i^{(n)}} - X_{t_{i-1}^{(n)}}\right)^2 \xrightarrow{p} \langle X\rangle_T$$

hold true if the mesh size $\delta_n = \sup_i \left(t_i^{(n)} - t_{i-1}^{(n)}, T - t_n^{(n)}, t_0^{(n)}\right)$ tends to zero as $n \to \infty$. These standard estimators are called realized covolatility and volatility or realized covariance and variance. We will use these names interchangeably.
The realized (co-)volatilities attain a $\delta_n^{-1/2}$-convergence rate in the sense that

$$\sum_{i=1}^n \left(X_{t_i^{(n)}} - X_{t_{i-1}^{(n)}}\right)\left(Y_{t_i^{(n)}} - Y_{t_{i-1}^{(n)}}\right) - \langle X, Y\rangle_T = \mathcal{O}_p\left(\delta_n^{1/2}\right)\ .$$

Beyond consistency and rates of convergence we put emphasis on the asymptotic distribution of estimators. A first detailed study of the asymptotic law of the realized volatility has been established in Jacod and Protter [2003]. As a side result of our general central limit theorem we can deduce that the estimation error of the realized covolatility multiplied with $\sqrt{n}$ converges stably in law to a mixed normal limiting distribution. In the Itô process model with observation noise, however, for some time instant $t_i^{(n)} - t_{i-1}^{(n)} = \mathcal{O}(n^{-1})$, the increments

$$X_{t_i^{(n)}} - X_{t_{i-1}^{(n)}} = \underbrace{\int_{t_{i-1}^{(n)}}^{t_i^{(n)}} \mu_t^X\, dt}_{=\mathcal{O}_p(n^{-1})} + \underbrace{\int_{t_{i-1}^{(n)}}^{t_i^{(n)}} \sigma_t^X\, dB_t^X}_{=\mathcal{O}_p\left(n^{-1/2}\right)} + \underbrace{\epsilon_{t_i^{(n)}}^X - \epsilon_{t_{i-1}^{(n)}}^X}_{=\mathcal{O}_p(1)}$$

are substantially governed by the observation errors. On the contrary, the influence of the drift terms is asymptotically small and the drift can be viewed as a nuisance term for the considered estimation problem. Due to the domination of noise corruption $\sqrt{n}$-consistent estimators cannot be achieved.

Groundwork and main findings

First of all, we grasp the corresponding estimation problem in a parametric world from which we can gain a better understanding about asymptotic optimality. In a univariate latent Itô process model with additive Gaussian noise, Gloter and Jacod [2001] have proved that for a constant volatility parameter σ, $n^{1/4}$ constitutes a lower bound for

the rate of convergence. An analogous result with the same rate could be established for the estimation of a constant correlation coefficient ρ in the synchronous bivariate setting which is one of the central results of this work. These findings are based on the fundamental concept by Le Cam [1960] about local asymptotic normality (LAN) of sequences of statistical models. This concept allows for a unified theory covering the wide variety of statistical models where the likelihood behaves locally and asymptotically like in a Gaussian shift model. By showing the LAN property, a lower bound for the rate of convergence is feasible by the convolution theorem and the local asymptotic minimax theorem of Hájek [1972]. Therefore, we are able to conclude that the rate $n^{1/4}$ can not be exceeded, also in the nonparametric framework. Furthermore, an asymptotic efficiency bound for the variances of sequences of estimators is obtained by the inverse Fisher information and the maximum-likelihood estimator (MLE) is known to be asymptotically efficient. Conversely, one can not deduce optimal rates from calculating the MLE.
During the last decade the nonparametric estimation problem of the quadratic variation in a latent Itô process model with microstructure noise has been studied intensively. This strand of literature followed Zhang et al. [2005] that has attracted a lot of attention to this estimation problem. Zhang et al. [2005] have constructed an estimator based on subsampling and a bias-correction and proved a stable central limit theorem with suboptimal $n^{1/6}$-rate. A refinement of the subsample approach using multiple scales in Zhang [2006] and related alternative techniques in Barndorff-Nielsen et al. [2008a], Podolskij and Vetter [2009] and Xiu [2010] have led to rate-optimal estimators and feasible stable central limit theorems. Yet, the estimation methods do not attain asymptotic efficiency. For the more specific nonparametric model with Gaussian noise Reiß [2011] has shown asymptotic equivalence in the Le Cam sense to a Gaussian shift experiment and could construct an asymptotically efficient estimator.
A methodology to deal with non-synchronous observations of Itô processes has been found by Hayashi and Yoshida [2005]. The so-called Hayashi-Yoshida estimator has superseded simpler previous-tick interpolation methods setting the standard for the estimation of the quadratic covariation from asynchronous observations in the absence of microstructure noise effects. The estimation approach that we propose for the most general case in the presence of noise and non-synchronicity arises as a combination of the multiscale estimator to handle noise contamination on the one hand and a synchronization algorithm in accordance with the Hayashi-Yoshida estimator to cope with non-synchronicity on the other hand. A first attempt in the same direction, combining one-scale subsampling and the Hayashi-Yoshida estimator, has been given in Palandri [2006]. We take up this synchronization method for our approach. An advance of the combined procedure and the progress to a multiscale estimator inspired by Zhang [2006], has improved upon existing methods and has led to the first rate-optimal estimator for the general setting. This generalized multiscale estimator has been introduced in Bibinger [2011] (first version 2008) in which, moreover, the LAN result has been published.
The main result of the work on hand is a feasible stable central limit theorem for the estimation error of the generalized multiscale estimator which constitutes at this stage, up to the author's knowledge, the only result of this nature.
The notion of stable weak convergence going back to Rényi [1963] is essential for our

asymptotic theory. Stable weak convergence $X_n \overset{st}{\rightsquigarrow} X$ is the weak convergence of (X_n, Z) to (X, Z) for every measurable bounded random variable Z. The limiting random variables in stable limit theorems are defined on extensions of the original underlying probability spaces. The reason for us to involve this concept of a stronger mode of weak convergence is that mixed normal limiting distributions are derived where asymptotic variances are themselves strictly positive random variables. Provided we have a consistent estimator V_n^2 for such a random asymptotic variance V^2 on hand, the stable central limit theorem $X_n \overset{st}{\rightsquigarrow} VZ$ with Z distributed according to a standard Gaussian law, yields the joint weak convergence $(X_n, V_n^2) \rightsquigarrow (VZ, V^2)$ and also $X_n/V_n \rightsquigarrow Z$ and hence allows to perform statistical inference providing tests or confidence intervals.
In the proofs of our limit theorems we will 'remove' the drifts in the sense that after a transformation to an equivalent martingale measure stable central limit theorems for Itô processes without drift are proved and, as illustrated in Mykland and Zhang [2009], stability of the weak convergence ensures that the asymptotic law holds true under the original measure. In this sense stable convergence is commutative with measure change.
Since we are concerned in this work with a topic on which vibrant research leads to permanent new contributions, several valuable publications have appeared during the elaboration of this thesis. Some of these inputs have influenced the advance of this work and we give credit to the authors at the respective positions and some proposed concurrent alternative methods and we give a comparison or comment on those at suitable points.
Barndorff-Nielsen et al. [2008b] proposed a kernel-based method with a previous-tick interpolation to so-called refresh times and established a stable central limit theorem with non-optimal $N^{1/5}$-rate for a multivariate non-synchronous design. Their estimator, furthermore, ensures that the estimated covariance matrix is positive semi-definite. Since this has set a new standard we draw a comparison to this approach at several stages of this work, in particular working out the differences and similarities of the synchronization methods in Section 3.1 and implementations of both in the simulation study in Section 6.2. Christensen et al. [2010] have stated a combination of pre-averaging and the Hayashi-Yoshida estimator that attains the optimal rate.
One recent alternative approach by Aït-Sahalia et al. [2010] arises as a combination of the univariate quasi-maximum-likelihood method by Xiu [2010], the polarization identity and a generalized synchronization scheme which is different from the Hayashi-Yoshida ansatz that we will use. For a sequence of times, for that at least one observation of each process lies between consecutive times, and the mesh size tends to zero, one observation is taken from each interval. This includes as a special case the refresh time approach of Barndorff-Nielsen et al. [2008b] to which we shall often compare our method. Mathematically, the estimator of Aït-Sahalia et al. [2010] is shown to attain the optimal rate.

A finance oriented motivation

We consider asymptotics in a high-frequency observations setting where on a fixed time interval $[0, T]$ the distances between observation times tend to zero. These kind of methods are congruously interesting for applications to high-frequency data. Those occur nowadays in various application areas as neuroscience and climatology. Foremost the link to finance applications is foregrounded and has stimulated an alliance of economists and statisticians to participate in the research on this field. Especially the estimation of integrated volatility and integrated covariance from high-frequency financial time series data has become an issue of great importance. Daily integrated (co-)volatilities from high-frequency intraday returns serve as a basis for risk management as well as portfolio optimization and hedging strategies.

Since the seminal work by Black and Scholes [1973] it has become standard to model log-prices of securities as Itô diffusions or, in the mean time, more general as semimartingale processes. Although trading takes place at discrete times those continuous-time models provide a convincing description of the dynamics of assets and allow for developments of powerful tools and convenient procedures for computing and trading strategies.

The last years have seen an enormous increase of the amount of trading activities for many liquid securities. Paradoxically, the availability of high-frequency data necessitated a new angle on financial modeling. In fact, for every semimartingale the realized (co-)volatilities converge in probability to the quadratic (co-)variations. The so-called signature plot on the left-hand side of Figure 0.1 visualizes the realized volatility of a financial time series, taken from the data of the application study in this work, for different frequencies. The number of observations included in the evaluation of the realized volatility decreases rightwards. We speak of high frequencies if the time instants between incorporated observations are small, hence, if the values on the x-axis in Figure 0.1 are small, and low frequencies if these values are big. The sparse-sampled estimators are quite robust in a certain domain of frequencies and it had been common practice to take such an estimate for an ad hoc chosen frequency before subsampling and alternative methods were introduced. For very high-frequencies instead the realized volatility explodes. This effect, reported in Brown [1990] among others, is ascribed to market microstructure frictions. Sources of this market microstructure noise are manifold. An important role plays the occurrence of bid-ask spreads. Aside from that transaction costs, strategic trading, limited market depths and discreteness of prices spread out the structure of the long-run dynamics that can be characterized by semimartingales.

The additive noise model reproduces the effects driven by the influences of the market microstructure. A sparse-sampled estimator is, however, not a satisfying solution since this means throwing away most of the available observations and is therefore an inefficient use of information. Modern estimation approaches as the multiscale estimator have solved this dilemma.

When realized covariances are calculated for fixed frequencies and a previous-tick interpolation is applied, the so-called Epps effect described in Epps [1979] appears. The phenomenon that the realized covariance tends to zero at the highest frequencies is due to non-synchronicity effects. For multivariate estimation strategies, apart from taking market microstructure noise into account, one has to accomplish a way to deal with asynchronous observation schemes.

Thus, methods as the one developed in the work on hand should be practicable and investigated for their utility in financial applications. We take account of that by performing

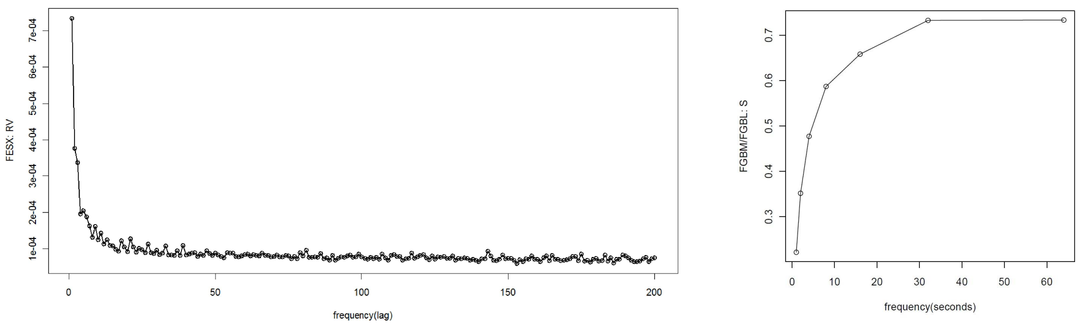


Figure 0.1: Signature plot and an illustration of the epps effect for high-frequency financial data.

an application study to future tick-data from the European Exchange (EUREX).

Summary with a preview of the key results

In Chapter 1 the theoretical foundation of the thesis is gathered. We set up the theoretical framework of stochastic calculus with an introduction to basic notation and an overview on fundamental results that we use in this work. A more broadly exposition is devoted to the essential concepts of Le Cam's local asymptotic normality, stable weak convergence and several central limit theorems for triangular arrays.

As a short guideline to the following chapters, note that Chapter 2 addresses noisy and synchronous, Chapter 3 non-noisy but asynchronous and Chapter 4 general noisy and asynchronous observations of latent Itô processes. The central Chapters 3 and 4 that come up with the proofs of stable central limit theorems are structured similarly for reasons of clarity. In each first section the estimation methods are illustrated and explained. The second section then provides an access to asymptotics and to deduce a stable central limit theorem in the sense that the main ingredients for the proof are illuminated. The detailed proofs are postponed to the third sections. Chapter 5 complements the theory of Chapter 4 for the estimation approach in the non-synchronous and noisy setting and Chapter 6 contains an application study. Note that assumptions which we only impose in certain chapters have a different numbering than those used throughout the whole work. First of all, in Section 2.1 we consider a closely related simple parametric model with two standard Brownian motions and a constant correlation coefficient ρ, that are observed at $(N+1)$ synchronous equidistant times with i. i. d. Gaussian noise. From this analysis we will derive a lower bound $N^{1/4}$ for the convergence rate of estimators of ρ that carries over to less informative situations as in the general nonparametric setting. Thanks to this result we can claim that the later developed estimator attains the optimal rate. For this purpose we show the LAN property with rate $N^{-1/4}$. This strategy has been inspired by Gloter and Jacod [2001] who established the corresponding one-dimensional result

for the estimation of a constant volatility parameter σ. The result can not be extended directly to the two-dimensional case for the correlation coefficient ρ. Nevertheless, parts of the proof follow the same principles and we give credit to Gloter and Jacod [2001] in the relevant passages.
The eigenvalues of the covariance matrix of the joint vector $(\tilde{X}, \tilde{Y})$ in (2.3a) and (2.3b) already give insight that one cannot hope to find an estimator with a faster rate than $N^{1/4}$. Only of order $\sqrt{N}$ of the first eigenvalues are dominated by the addend with the parameter of interest whereas for the others the noise terms are leading. This is in accordance with the finding that the increments are dominated by the impact of the noise. In subsequent calculations it is shown that the log-likelihood has locally and asymptotically the shape as for a Gaussian shift experiment constituting the LAN property in Theorem 2.1.
As a side result, we derive the bounds (2.1) for the asymptotic Fisher information

$$\frac{1}{8\eta_X}\left(\frac{1}{(1+\rho)^{3/2}}+\frac{1}{(1-\rho)^{3/2}}\right)\leq I(\rho)\leq\frac{\sqrt{2}}{8}\frac{1}{\sqrt{\eta_X^2+\eta_Y^2}}\left(\frac{1}{(1+\rho)^{3/2}}+\frac{1}{(1-\rho)^{3/2}}\right)$$

that provide a benchmark for the asymptotic variances of estimators for the quadratic covariation, where the dependence of the Fisher information on the correlation coefficient is of particular interest. The noise variances are denoted by η_X^2 and η_Y^2. If they are equal the upper equals the lower bound and we end up with the exact result (2.2) below.

In Section 2.2 the subsampling approach of Zhang et al. [2005] and the rate-optimal multiscale version by Zhang [2006] are extended to a bivariate synchronous design. The starting point is a one-scale subsample estimator

$$\widehat{\langle X,Y\rangle}_T^{sub}=\frac{1}{i}\sum_{j=i}^{n}\left(\tilde{X}_{t_j}-\tilde{X}_{t_{j-i}}\right)\left(\tilde{Y}_{t_j}-\tilde{Y}_{t_{j-i}}\right)\ .$$

that is motivated from two perspectives. The one in line with Zhang et al. [2005] is to (post-)average sparse-sampled lower frequent realized covariances and one is to evaluate a usual realized covariance from the time series on that we have run a linear filter first. The latter means that on a moving window noisy observations are (pre-)averaged first. This estimator corresponds to the univariate "second-best approach" in Zhang et al. [2005], but on the assumption of mutually independent microstructure processes a bias-correction that completed the "first-best approach" is redundant here. The bivariate multiscale estimator for synchronous sampling in (2.11a) has also the analogous form to its univariate origin.
Section 2.3 is devoted to two alternative approaches. One estimator by Barndorff-Nielsen et al. [2008a] to handle noise contamination arises as linear combination of autocovariances and is called the kernel-based approach. The other one is a pre-average method introduced in Podolskij and Vetter [2009]. As a reaction to the progress in research during the work on this thesis, the overview goes beyond these one-dimensional considerations and summarizes also advancements of the methods up to the current stage that also gear

towards more broadly settings, particularly noisy and non-synchronous observations. Our choice to build up an estimator on the combination of a synchronization technique with the multiscale estimator is not unique. Since very close relations between the three approaches to tackle noise corruption have been revealed, it is plausible that an estimation approach in the presence of asynchronicity and microstructure noise is possible by combinations with any of those methods.

In Chapter 3 we focus on the problem of estimating the quadratic covariation from non-synchronously observed Itô processes X and Y at times $t_i^{(n)}$ and $\tau_j^{(m)}$ with $m \sim n$ meaning that $m = \mathcal{O}(n)$ and $n = \mathcal{O}(m)$. This problem has been solved in Hayashi and Yoshida [2005] in the sense that a $\sqrt{n}$-consistent estimator has been found. On further regularity assumptions asymptotic normality of the estimator has been proved in Hayashi and Yoshida [2008] for deterministic volatility and correlation functions.
Nevertheless, it will turn out to be convenient as first stage of our combined approach to reinvent the Hayashi-Yoshida estimator

$$\widehat{\langle X, Y\rangle}_T^{(HY)} = \sum_{i=1}^{n}\sum_{j=1}^{m} \Delta X_{t_i} \Delta Y_{\tau_j} \mathbb{1}_{[\min(t_i,\tau_j) > \max(t_{i-1},\tau_{j-1})]}$$

in a slightly different manner. The estimator that is the sum of all products of increments $\Delta X_{t_i} = X_{t_i} - X_{t_{i-1}}$ and $\Delta Y_{\tau_j} = Y_{\tau_j} - Y_{\tau_{j-1}}$ with overlapping observation time instants can be rewritten using previous and next-tick interpolations:

$$\begin{aligned}\widehat{\langle X, Y\rangle}_T^{(HY)} &= \sum_{i=1}^{n} \Delta X_{t_i}\left(Y_{t_{i,+}} - Y_{t_{i-1,-}}\right)\\ &= \sum_{i=1}^{N} (X_{g_i} - X_{l_i})(Y_{\gamma_i} - Y_{\lambda_i}) = \sum_{i=1}^{N} (X_{T_{i,+}^X} - X_{T_{i-1,-}^X})(Y_{T_{i,+}^Y} - Y_{T_{i-1,-}^Y})\,,\end{aligned}$$

where $t_{i,+} := \min_{0 \le j \le m}(\tau_j | \tau_j \ge t_i)$ and $t_{i,-} := \max_{0 \le j \le m}(\tau_j | \tau_j \le t_i)$ and analogously below. The first illustration, serving also as a good implementation rule, can as well be written in the symmetric way. The second rewriting relies on the 'translation' of the principle of the Hayashi-Yoshida estimator into the iterative synchronization Algorithm 3.1 adopted from Palandri [2006]. The last rewriting above hints already at our ansatz to establish the asymptotic theory. The T_is, $i = 0, \ldots, N$, are defined by a partition of $[0, T]$ that we call the closest synchronous approximation and N corresponds to the number of constructed sets by our algorithm. It turns out that the T_is are exactly the refresh times from Barndorff-Nielsen et al. [2008b]. The difference of the two synchronization methods is hence the replacement of previous-tick by next-tick interpolation at right end points of instants $(T_i - T_{i-1})$ in accordance with the Hayashi-Yoshida estimator. Next-tick interpolations are always feasible for these kind of ex-post estimation problems. The overall estimation error ignoring boundary terms can be split in a 'familiar' synchronous type discretization error D_T^N from (3.5) and an asymptotically independent error due to the lack of synchronicity A_T^N from (3.6). We apply a stable convergence theorem from Jacod [1997] which provides a suitable concept for the setting considered in this work. It allows us to prove stable weak convergence of stochastic processes associated with D_t^N and A_t^N, for $t \in [0, T]$, to limiting time-changed Brownian motions. The stable weak convergence of $A_T^N + D_T^N$ to a centred mixed normal limit is implied as the marginal distribution for $t = T$. The convergence of the sequences of variances to asymptotic variances hold if the following quadratic variation of times

$$G^N(t) = \frac{N}{T} \sum_{T_i^{(N)} \le t} \left(\Delta T_i^{(N)}\right)^2 ,$$

and certain covariations of times F^N and H^N given in (3.7b) and (3.7c) that hinge on interpolation steps, converge to continuously differentiable limiting functions and the sequences of difference quotients converge uniformly. The asymptotic quadratic variation of time G of the $T_i^{(N)}$s influences the asymptotics of D_T^N. The covariation of times F^N measures an interaction of interpolation errors between the two processes and H^N the impact of the in general non-zero correlations of the products involving previous- and next-tick interpolations at the same $T_i^{(N)}$s for each process separately. The limiting functions F and H contribute to the asymptotic shape of A_T^N. These convergence assumptions seem to be rather mild and weaker than an assertion that the joint sampling schemes design has to tend to some limiting design with a certain asymptotic behaviour of asynchronicity. Time-homogeneous observation schemes lead to linear limiting functions on $[0, T]$. On the assumption that there exists a constant $\alpha > 0$ such that $\sup_i \Delta T_i^{(N)} = \mathcal{O}\left(n^{-2/3-\alpha}\right)$ and the Novikov condition to apply Girsanov's theorem, we conclude

$$\sqrt{N}\left(\sum_{i=0}^{N} \left(X_{g_i} - X_{l_i}\right)\left(Y_{\gamma_i} - Y_{\lambda_i}\right) - \langle X\,,\,Y\rangle_T\right) \stackrel{st}{\rightsquigarrow} \mathbf{N}\left(0\,,\,v_{D_T} + v_{A_T}\right) ,$$

with the asymptotic variance

$$T\int_0^T G'(t)\left(\sigma_t^X\sigma_t^Y\right)^2\left(\rho_t^2+1\right)dt + T\int_0^T \left(F'(t)\left(\sigma_t^X\sigma_t^Y\right)^2 + 2H'(t)\left(\rho_t\sigma_t^X\sigma_t^Y\right)^2\right)dt\,.$$

This main result of Chapter 3 in Theorem 3.1 has improved upon the asymptotic normality result in Hayashi and Yoshida [2008] since the weak convergence is stable and holds for random volatility and correlation processes. Independently, Hayashi and Yoshida [2011] have proved a result of the same kind. Yet, the above limit theorem and the differing ansatz to the asymptotic analysis in this chapter is valuable to further elucidate how non-synchronicity affects the asymptotics. Most of all it serves as a good preparation for the construction of the general approach in Chapter 4.

The original new estimation approach introduced in Chapter 4 copes with the impact of the noise contamination and asynchronous observation schemes. The microstructure noise processes ϵ^X and ϵ^Y are assumed to be independent of X and Y. The so-called generalized multiscale estimator

$$\widehat{\langle X,Y\rangle}_T^{multi} = \sum_{i=1}^{M_N} \frac{\alpha_{i,M_N}^{opt}}{i} \sum_{j=i}^{N} \left(\tilde{X}_{g_j^{(N)}} - \tilde{X}_{l_{j-i+1}^{(N)}} \right) \left(\tilde{Y}_{\gamma_j^{(N)}} - \tilde{Y}_{\lambda_{j-i+1}^{(N)}} \right) ,$$

and its one-scale version (4.3) have thanks to our synchronization practice an appearance which is very close to the estimators for synchronous sampling. The idea is to apply the subsampling methods to a fictive idealized synchronized design with observations at the T_is. Instead of averaging sparse-sampled Hayashi-Yoshida type estimators, what can also lead to a rate-optimal multiscale type estimator but with larger variance, we perform only interpolations on the highest frequency scale at right and left end points of low frequent increments. An important advantage of this procedure is that the covariations of times do not affect the asymptotics any more. There is a trade-off between the variance due to noise, which is of order N/M_N^3, and the discretization variance of order M_N/N. Cross terms and an error term due to the combination of noise and boundary effects are of order $1/\sqrt{M_N}$. Hence, choosing $M_N = c_{multi}\sqrt{N}$ minimizes the overall mean square error and we attain the optimal rate $N^{1/4}$. The same weights as in Zhang [2006] for the univariate estimator are incorporated, since they also solve the minimization problem of the error due to noise with the side condition of asymptotic unbiasedness here. The error due to the lack of synchronicity is asymptotically negligible for the total discretization error but still non-synchronicity will have an effect on the asymptotic variance. In particular, the interpolations demand that certain observations of $\tilde{X}$ and $\tilde{Y}$ appear twice in the estimator. In the microstructure noise setting this means that the corresponding errors are involved twice. The errors due to noise and cross terms hinge on the number of such events when carrying out the synchronization. In Section 4.2 all possible aggregations in the sampling schemes are disentangled. We express the described effect by introducing degrees of regularity of non-synchronicity $I_X^N(t)$ and $I_Y^N(t)$ defined in Definition 4.2.1. On an analogous convergence assumption as for the covariations of times above, the Novikov condition and if there is a constant $\alpha > 0$ such that $\sup_i \Delta T_i^{(N)} = \mathcal{O}\left(n^{-8/9-\alpha}\right)$, with the theory from Jacod [1997], we can prove that:

$$N^{1/4}\left(\widehat{\langle X,Y\rangle}_T^{multi} - \langle X,Y\rangle_T\right) \stackrel{st}{\rightsquigarrow} \mathbf{N}\left(0,\mathbf{AVAR}_{multi}\right)$$

with the asymptotic variance

$$\begin{aligned}\mathbf{AVAR}_{multi} = c_{multi}^{-3}&\left(24 + 12\,\frac{I_X(T)+I_Y(T)}{T}\right)\eta_X^2\eta_Y^2 + c_{multi}^{-1}\,\frac{12\eta_X^2\eta_Y^2}{5}\\ &+ c_{multi}\,\frac{26}{35}T\int_0^T G'(t)(\sigma_t^X\sigma_t^Y)^2(1+\rho_t^2)\,dt\\ &+ c_{multi}^{-1}\,\frac{12}{5}\left(\eta_Y^2\int_0^T (1+I_Y'(t))(\sigma_t^X)^2\,dt + \eta_X^2\int_0^T (1+I_X'(t))(\sigma_t^Y)^2\,dt\right.\end{aligned}$$

This main result of Theorem 4.1 is proved on the assumption that the noise processes are mutually independent, i. i. d. and that the expectation equals zero and fourth moments are finite. As a side result we also obtain a stable central limit theorem as Corollary 4.2.2 for the one-scale estimator. In a completely asynchronous setting $I^X \equiv I^Y$ holds.

Section 5.1 comes up with a consistent estimation of the asymptotic variances that appear in the stable limit theorems that lay the foundation to make use of the concept of stable convergence for statistical inference and makes the asymptotic mixed normality result feasible. The estimator consist of an adopted estimator $(1/2n)\sum(\Delta X_{t_i})^2$ of the noise variance η_X^2 and analogously for η_Y^2 and histogram-based estimators of the integrals that appear. For their construction bins are chosen equispaced according to the timelines associated with the respective functions whose derivatives occur in the integral. Then on each bin multiscale estimators in the noisy, and Hayashi-Yoshida type estimators in the asynchronous non-noisy case, are evaluated to estimate the local quadratic (co-)variations. In Section 5.2 we are concerned with mutually independent homogeneous Poisson sampling schemes. The theory for deterministic sampling includes random sampling according to some process independent of $\tilde{X}$ and $\tilde{Y}$ when regarding the conditional law. We can explicitly determine the functions F, H, G, I_X and I_Y for arbitrary parameters θ_1 and θ_2 of the Poisson processes where the convergence is in probability and give special versions of the limit theorems from Chapter 3 in Proposition 5.2.2 and 4 in Proposition 5.2.3 where the asymptotic variance yields

$$
\begin{aligned}
\mathbf{AVAR}_{multi}^{poiss} &= c_{multi}^{-3}\left(24+12\,\frac{2\theta_1\theta_2}{(\theta_1+\theta_2)^2}\right)\eta_X^2\eta_Y^2 + c_{multi}^{-1}\,\frac{12\eta_X^2\eta_Y^2}{5}\\
&\quad + c_{multi}\,\frac{26}{35}\int_0^T 2\left(1-\frac{2\theta_1^2\theta_2^2}{\theta_1^2\theta_2^2+(\theta_1^2+\theta_2^2)(\theta_1+\theta_2)^2}\right)(\sigma_t^X\sigma_t^Y)^2(1+\rho_t^2)\,dt\\
&\quad + c_{multi}^{-1}\,\frac{12}{5}\left(\eta_Y^2\int_0^T(1+\frac{\theta_1\theta_2}{(\theta_1+\theta_2)})(\sigma_t^X)^2\,dt\;+\eta_X^2\int_0^T(1+\frac{\theta_1\theta_2}{(\theta_1+\theta_2)})(\sigma_t^Y)^2\,dt\right)\,.
\end{aligned}
$$

There are several important modifications of the model and some assumptions can be relaxed. For the use in financial applications one has to encounter the fact that an i. i. d. noise is often unrealistic. Market microstructure frictions are mainly induced by bid-ask spreads and there is a tendency that alternating buy– and sell–market orders, which are not committed to certain strike prices as opposed to limit orders, drive the observation noise leading to the structure of negative correlations between succeeding trades. We show that the generalized multiscale estimator can cope with serial dependence in the noise as long as mixing coefficients decay exponentially and remains consistent, asymptotically unbiased and rate-optimal. Though, the asymptotic variance increases and a closed-form expression is in general not feasible. This and reasons of clarity and comprehensibility motivated us to first carry out the theory for a more restrictive i. i. d. assertion. The theory for serial dependent microstructure noise has been developed in detail for the univariate setting in Aït-Sahalia et al. [2009].
An additional point is that empirical studies suggest to rather model the noise variance to decrease with the number of observations. An amiable feature of the generalized

multiscale estimator is that the Hayashi-Yoshida nature of its synchronization inherits a bridge between the noisy and the non-noisy case. For decreasing noise variance $\eta^2_{X,N}, \eta^2_{Y,N} \sim N^{-\alpha}$ for some $0 < \alpha < 1$, the rate of the generalized multiscale estimator improves to $N^{\frac{1}{4}+\frac{\alpha}{4}}$. This is a clear advantage over methods that use only previous-tick interpolations. An interesting question is if and how one can include a more general semimartingale process allowing for jumps as efficient process in the model. We offer one possible practice relying on Fan and Wang [2007] to do so.

For an implementation of the estimator and its use in applications, the multiscale frequency $M_N = c_{multi}\sqrt{N}$ has to be chosen first. As many other nonparametric estimation techniques the estimator hinges on a tuning parameter. In Section 6.1 we state a convenient algorithm to derive an accurate choice. For this purpose we calculate the histogram-based estimators for the terms appearing in the asymptotic variance first. The tuning parameters of those are determined adaptively by pilot estimates. We use the resulting estimates to estimate that constant c_{multi} that minimizes the asymptotic variance. It turns out in the simulation part in Section 6.2 that the estimators are quite robust to the involved frequencies and that our algorithm provides adequate choices. The other main findings of the application study are that our method outperforms the one of Barndorff-Nielsen et al. [2008b] for mild noise levels whereas both perform well, and better than the asymptotically less efficient one-scale version, for very high noise levels. The empirical application to EUREX tick-data in Section 6.3 contains also tests for hypotheses that integrated covariances are zero for which we have asymptotic distribution free tests directly on hand from the feasible stable central limit theorems. Those reveal that we can reject zero covariations between two German federal bonds and two related stock indices with very small p-values. Estimated noise levels are quite small. Thus, it is an advantage to use our synchronization approach instead of pure previous-tick interpolations.

1 Theoretical Concepts

1.1 A concise survey of important theorems from stochastic calculus

1.1.1 Stochastic integration and quadratic (co-)variation

The following stochastic processes are defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ with a right-continuous ($\mathcal{F}_t = \bigcap_{u>t} \mathcal{F}_u \ \forall t \geq 0$) and complete ($\mathcal{F}_0$ contains all $\mathbb{P}$-null sets) filtration $(\mathcal{F}_t)_{0 \leq t \leq \infty}$. A process $(H_t)_{0 \leq t \leq \infty}$ is called simple predictable with respect to $(\mathcal{F}_t)_{0 \leq t \leq \infty}$, if it has a representation

$$H_t = \xi_0 \mathbb{1}_0(t) + \sum_{i=1}^{n} \xi_i \mathbb{1}_{(T_i, T_{i+1}]}(t) \ ,$$

where $0 = T_1 \leq \ldots \leq T_{n+1} < \infty$ is a finite sequence of stopping times and ξ_i are $(\mathcal{F}_{T_i})$-measurable almost surely finite random variables. The set of simple predictable processes topologized by uniform convergence in (t, ω) is denoted $\mathbf{S}$. The space of finite-valued random variables topologized by convergence in probability is denoted $\mathbb{L}^0(\mathbb{P})$. For a stochastic process X and fixed t the stochastic integral of simple predictable processes with respect to X is defined by the linear mapping

$$\mathbf{I}_X : \mathbf{S} \longrightarrow \mathbb{L}^0(\mathbb{P}), \ H_t = \xi_0 \mathbb{1}_0(t) + \sum_{i=1}^{n} \xi_i \mathbb{1}_{(T_i, T_{i+1}]}(t) \longmapsto \xi_0 X_0 + \sum_{i=1}^{n} \xi_i \left(X_{T_{i+1}} - X_{T_i} \right) \ .$$

$\mathbf{I}_X$ does not depend on the choice of the representation of H in $\mathbf{S}$.

Definition 1.1.1. *An adapted càdlàg process X is called a total semimartingale if the mapping $\mathbf{I}_X : \mathbf{S} \longrightarrow \mathbb{L}^0(\mathbb{P})$ is continuous. X is called a semimartingale if, for all $t \in [0, \infty)$, the process $(X_{\tau \wedge t})_{\tau \geq 0}$ is a total semimartingale.*

Note that we use angle instead of square brackets in the following definition unlike the commonly used notation in the literature.

Definition 1.1.2. *The quadratic variation process of a semimartingale X is defined by*

$$\langle X \rangle = X^2 - 2 \int X_- \, dX, \quad \text{where} \quad X_- := \lim_{u \to s, u < s} X_u, \ (X_-)_0 = 0 \ .$$

The quadratic covariation process of two semimartingales X and Y is defined by

$$\langle X, Y\rangle = X \cdot Y - \int X_- \, dY - \int Y_- \, dX \ .$$

The last preceding basic definitions are taken from the book by Protter [2004]. We use the different notation for the quadratic (co-)variation, since we will not consider the conditional quadratic (co-) variation process which is the compensator of the quadratic (co-)variation process and which exists when the quadratic (co-)variation is locally bounded. For continuous semimartingales that we will focus on throughout this work both processes are equal.

Proposition 1.1.3. *Let $X, Y, Z, \tilde{Z}$ be semimartingales. It holds true that:*

- *$\langle X\rangle$ is an adapted non-decreasing process with càdlàg paths of finite variation.*
- *The mapping $(X, Y) \longrightarrow \langle X, Y\rangle$ is symmetric and bilinear.*
- *$d(X_t \cdot Y_t) = (X_-)_t \, dY_t + (Y_-)_t \, dX_t + d\langle X, Y\rangle_t$ (integration by parts)*
- *$\langle \int Z \, dX, \int \tilde{Z} \, dY\rangle_t = \int_0^t Z_\tau \tilde{Z}_\tau \, d\langle X, Y\rangle_\tau$ for $Z, \tilde{Z}$ càdlàg.*
- *For partitions $\Pi = \{0 = t_0, ..., t_n = t\}$ of $[0, t]$:*
$$\sum_{k=1}^{n} \left(X_{t_i} - X_{t_{i-1}}\right)^2 \xrightarrow{p} \langle X\rangle_t \quad as \quad ||\Pi|| := \max_{1\le l\le n} \left(|t_l - t_{l-1}|\right) \to 0 \, .$$

The last statement is proved as Theorem 22 in Section II.6 in Protter [2004], the integration by parts formula in Corollary 2 of Section II.6. The fourth statement is included in Theorem 29. The other points are well-known and we refer to literature on stochastic calculus for more information (e. g. Revuz and Yor [1991], Shreve [2008], Karatzas and Shreve [1991]).
An adapted càdlàg process M is a local martingale if a sequence $T_n \uparrow \infty$ (a. s.) of stopping times exists such that $M_{t\wedge T_n} \mathbb{1}_{\{T_n>0\}}$ is a uniformly integrable martingale for each n.

Proposition 1.1.4 (Burkholder-Davis-Gundy inequalities). *Let M be a continuous local martingale with $M_0 = 0$, then for every $p > 0$ there exist constants c_p and C_p, such that*

$$c_p \, \mathbb{E}\left[\langle M, M\rangle_T^{p/2}\right] \le \mathbb{E}\left[(M_T^*)^p\right] \le C_p \, \mathbb{E}\left[\langle M, M\rangle_T^{p/2}\right]$$

for every stopping time T holds true, where $M_t^* = \sup_{s\le t} |M_s|$.

The constants appearing in the Burkholder-Davis-Gundy inequalities (BDG) are universal in the sense that they only depend on p, but not on M and the underlying probability space. In particular, if
$\mathbb{E}\left[\sqrt{\langle M, M\rangle_t}\right] < \infty, \ \forall \ 0 < t < \infty$ holds, M is a martingale. A proof of the BDG-inequalities is given in Revuz and Yor [1991], Section IV.4.

Throughout this work we are mainly concerned with Itô processes of the type $X_t = \int_0^t \mu_s \, ds + \int_0^t \sigma_s \, dB_s$ with a Brownian motion B_s and continuous processes σ_s and locally bounded μ_s. From Proposition 1.1.3 we know that the quadratic variation of such an Itô process $\langle \int \sigma_s \, dB_s, \int \sigma_s \, dB_s \rangle_t = \int_0^t \sigma_s^2 \, ds$ is the so-called integrated volatility (or integrated variance) since the drift part is of finite variation. All processes of finite variation have a quadratic variation equal to zero. For the quadratic covariation of two Itô processes with correlated Brownian motions such that $d\langle B^X, B^Y \rangle_s = \rho_s$ with a continuous process ρ_s, $\langle X, Y \rangle_t = \int_0^t \rho_s \sigma_s^X \sigma_s^Y ds$ is obtained.

Theorem 1.1 (**Itô's lemma**). *Let X be a continuous semimartingale and $f \in \mathcal{C}^2(\mathbb{R})$. Then $f(X)$ is again a semimartingale and the following Itô formula holds true:*

$$f(X_t) = f(X_0) + \int_0^t f'(X_s) \, dX_s + \frac{1}{2} \int_0^t f''(X_s) \, d\langle X \rangle_s \; .$$

A proof of Itô's lemma can be found in Protter [2004] (Theorem II.32).
For processes of finite variation, we state the following Corollary:

Corollary 1.1.5 (**Change of variables**). *Let A be a process with continuous paths of finite variation and $f \in \mathcal{C}^1(\mathbb{R})$. Then $f(A)$ is a process of finite variation and it holds true that:*

$$f(A_t) = f(A_0) + \int_0^t f'(A_s) \, dA_s \; .$$

For a process A having right-continuous paths of finite variation the following generalization holds true:

$$f(A_t) = f(A_0) + \int_0^t f'(A_{s-}) \, dA_s + \sum_{0<s\leq t} \left(f(A_s) - f(A_{s-}) - f'(A_{s-}) \Delta A_s \right) \; .$$

We refer to Theorem 31 and Theorem 34 of Chapter 2 in Protter [2004] for the proofs. The Corollary is based on the fact, that for a finite variation process A with continuous paths and a càdlàg process H, the stochastic integral $\int H_s \, dA_s$ is indistinguishable from the Lebesgue-Stieltjes integral computed path-by-path (Theorem II. 17 in Protter [2004]).

1.1.2 Girsanov's theorem

Let $B = \left\{ B_t = \left(B_t^{(1)}, ..., B_t^{(d)} \right), \mathcal{F}_t \right\}$ be a d-dimensional Brownian motion and $X = \{X_t, \mathcal{F}_t\}$ a vector of measurable adapted processes with

$$\int_0^T \left(X_t^{(i)} \right)^2 dt < \infty \;\; a.s. \;\; \forall i \in \{1, ..., d\}, \; 0 \leq T < \infty \; .$$

The process defined by

$$Z_t(X) := \exp\left(\sum_{i=1}^{d}\int_0^t X_s^{(i)}\ dB_s^{(i)} - \frac{1}{2}\int_0^t ||X_s||^2\ ds\right)$$

is a continuous local martingale. If Z is a martingale, for all $0 \leq T < \infty$, $\tilde{\mathbb{P}}_T(A) = \mathbb{E}\left[\mathbb{1}_A Z_T(X)\right]$
$\forall A \in \mathcal{F}_T$ defines a probability measure $\tilde{\mathbb{P}}_T$ on $\mathcal{F}_T$.

Theorem 1.2 (Girsanov). *If $Z(X)$ is a martingale, the process $\left(\tilde{B}, \mathcal{F}_t, 0 \leq t < T\right)$ defined by*

$$\tilde{B}_t^{(i)} = B_t^{(i)} - \int_0^t X_s^{(i)}\ ds, \quad i \in \{1, ..., d\} \quad 0 \leq t < T$$

is a d-dimensional Brownian motion on $(\Omega, \mathcal{F}_T, \tilde{\mathbb{P}}_T)$ for every fixed $T \in [0, \infty)$.

This version of Girsanov's theorem is proved in Karatzas and Shreve [1991], page 191 ff. A sufficient criterion to verify that $Z(X)$ is a martingale is Novikov's criterion $\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T ||X_s||^2\ ds\right)\right] < \infty$, $0 \leq T < \infty$.
Consider an Itô process X as above with $\sigma_t > 0$ being strictly positive. There is an adapted process γ_s with $\int_0^T \gamma_s^2\ ds < \infty$ *a. s.*, $0 \leq T < \infty$ and $\sigma_s\gamma_s + \mu_s = 0$. Assume $\mathbb{E}\left[\exp\left((1/2)\int_0^T \gamma_s^2\, ds\right)\right] < \infty$. By Girsanov's theorem there is a measure $\tilde{\mathbb{P}}$, equivalent to $\mathbb{P}$, with

$$\mathrm{d}\mathbb{P}/\mathrm{d}\tilde{\mathbb{P}} = \exp\left(-\int_0^t \gamma_s dB_s + \frac{1}{2}\int_0^t \gamma_s^2\ ds\right), \tag{1.1}$$

where $\mathrm{d}\mathbb{P}/\mathrm{d}\tilde{\mathbb{P}}$ is a $\mathbb{P}$-martingale and $\tilde{B}_t := B_t - \int_0^t \gamma_s\ ds$ is a $\tilde{\mathbb{P}}$-Brownian motion. The process X has, after a change of measure to $\tilde{\mathbb{P}}$, a martingale-representation. $\tilde{\mathbb{P}}$ is called the equivalent martingale measure to $\mathbb{P}$. For X we can now write

$$dX_t = \mu_t dt + \sigma_t dB_t = \sigma_t d\tilde{B}_t\ .$$

For non-random bounded μ_t and σ_t the conditions that $\sigma_t > 0$ is non-degenerate and $|\mu_t|/(\sigma_t)^2$ locally integrable are sufficient. This is proved in Karatzas and Shreve [1991], page 339 ff. For further specific cases and the assumptions on which the Girsanov transformation is possible we refer to Lipster and Shiryaev [2001].
We will apply Girsanov's theorem to prove results about weak convergence and convergence in probability assuming without loss of generality that the drift is equal to zero ($\mu_t \equiv 0$). Since if, for fixed T, the results hold under the equivalent martingale measure $\tilde{\mathbb{P}}$, they also hold true under $\mathbb{P}$. In particular, if we find a consistent estimator $\hat{\theta}_n$ for a parameter θ and can prove an asymptotic normality result that $n^{\alpha}(\hat{\theta}_n - \theta) \rightsquigarrow m + \mathbf{AVAR} \cdot Z$ with $Z \sim \mathbf{N}(0, 1)$ under $\tilde{\mathbb{P}}$, consistency and the rate of convergence will also be true under the equivalent measure $\mathbb{P}$. However, the asymptotic law under $\mathbb{P}$ may differ from the

asymptotic distribution under $\tilde{\mathbb{P}}$. We only know that the normal distribution remains after a change of measures. In the next section, we will introduce the concept of stable convergence in law which will, furthermore, guarantee that the asymptotic distributions of the limiting random variable are equal under $\mathbb{P}$ and $\tilde{\mathbb{P}}$ (see Mykland and Zhang [2009]). We also refer to Revuz and Yor [1991], page 327 ff., and Section 8.6 in Øksendal [2005] for further information concerning the change of measures.

1.1.3 The time-change theorem

In this subsection the theorem of Dambis and Dubins-Schwarz is presented which gives a fundamental connection between continuous local martingales and the Brownian motion. In particular, every continuous local martingale has a representation as a time-changed Brownian motion which emphasizes the key role of Brownian motion in the continuous martingale theory.
Let A_t be a non-decreasing right-continuous process, adapted to a right-continuous filtration $(\mathcal{F}_t)$.

$$T_\tau = \inf\{t : A_t > \tau\}, \quad \inf\{\emptyset\} := +\infty$$

defines a non-decreasing right-continuous family of stopping times. T_τ regarded as function of τ is called the right-continuous inverse of A. For all τ the random variable A_t is a $(\mathcal{F}_{T_\tau})$-stopping time, since $A_t = \inf\{\tau : T_\tau > t\}$.

Definition 1.1.6. *A sequence of stopping times T_τ is called a time-change, if the mapping $\tau \mapsto T_\tau$ is almost surely non-decreasing and right-continuous.*

Theorem 1.3 (**Dambis, Dubins-Schwarz**). *Let M be a $(\mathcal{F}_\tau, \mathbb{P})$-continuous local martingale, $M_0 = 0$ and $\langle M, M\rangle_\infty = \infty$, and T_τ the time-change*

$$T_\tau = \inf\{s : \langle M, M\rangle_s > \tau\} \ .$$

Then $B_\tau = M_{T_\tau}$ is a $(\mathcal{F}_{T_\tau})$-Brownian motion and $M_\tau = B_{\langle M,M\rangle_\tau}$. This Brownian motion B is called Dambis, Dubins-Schwarz (DDS) Brownian motion of M.

For the proof it can be shown that B is a continuous local $(\mathcal{F}_{T_\tau})$-martingale and $\langle B, B\rangle_\tau = \langle M, M\rangle_{T_\tau} = \tau$. The proof that it is a Brownian motion is based on Lévy's theorem that standard Brownian motion is the only continuous local martingale with $\langle B\rangle_t = t$ (cf. Theorem II.39 in Protter [2004]). Since $B_{\langle M,M\rangle} = M_{T_{\langle M,M\rangle}}$ and $M_{T_{\langle M,M\rangle_\tau}} = M_\tau$, one concludes that $B_{\langle M,M\rangle} = M$. We refer to Theorem V.1.6 in Revuz and Yor [1991] for a thorough proof of the theorem.
The restriction to $\langle M, M\rangle_\infty = \infty$ ensures that T_τ is almost surely finite and the DDS-Brownian motion can be defined on $\left(\Omega, \tilde{\mathcal{F}}_\tau, \mathbb{P}\right)$ with $\tilde{\mathcal{F}}_\tau = \mathcal{F}_{T_\tau}$. If $\langle M, M\rangle_\infty < \infty$, M has a representation as a time-changed Brownian motion on a suitable augmentation of the underlying probability space.

Definition 1.1.7. *An enlargement of the filtered probability space* $(\Omega, \mathcal{F}_\tau, \mathbb{P})$ *is another probability space* $(\tilde{\Omega}, \tilde{\mathcal{F}}_\tau, \tilde{\mathbb{P}})$ *and a mapping* $\sigma : \tilde{\Omega} \longrightarrow \Omega$ *with* $\sigma^{-1}(\mathcal{F}_\tau) \subset \tilde{\mathcal{F}}_\tau \ \forall \tau$ *and* $\sigma(\tilde{\mathbb{P}}) = \mathbb{P}$*. The latter means that* $\tilde{\mathbb{P}}\left(\sigma^{-1}(A)\right) = \mathbb{P}(A)$ *for all* $A \in \tilde{\mathcal{F}}$*.*

A process X defined on Ω may be viewed as defined on $\tilde{\Omega}$ by setting $X(\tilde{\omega}) = X(\omega)$ if $\sigma(\tilde{\omega}) = \omega$.

Proposition 1.1.8. *Assume all assumptions of Theorem 1.3, but* $\langle M, M\rangle_\infty < \infty$*. There is an enlargement* $(\tilde{\Omega}, \tilde{\mathcal{F}}_\tau, \tilde{\mathbb{P}})$ *of* $(\Omega, \mathcal{F}_\tau, \mathbb{P})$ *and a Brownian motion* B^* *defined on* $\tilde{\Omega}$*, independent of* M*, such that*

$$B_\tau = \begin{cases} M_{T_\tau} & \text{if } \tau < \langle M, M\rangle_\infty \\ M_\infty + B^*_{\tau - \langle M,M\rangle_\infty} & \text{if } \tau \geq \langle M, M\rangle_\infty \end{cases}$$

is a standard Brownian motion.

The theorem and the proposition can also be extended to the d-dimensional setting (Knight-Theorem). We refer to Section V.1 of Revuz and Yor [1991] for the proofs and further results on time-changed martingales. Section VIII.2 of the same book contains results on asymptotics and weak convergence of time-changed Brownian motions. We restrict ourselves to the following Theorem 1.4 which we call asymptotic Knight theorem:

Theorem 1.4 (**asymptotic Knight-theorem**)**.** *Let* M_n *be a sequence of continuous local martingales,* $(M_n)_0 = 0$ *and* $\langle M_n, M_n\rangle_\infty = \infty$*. Let* T^n_τ *be the time-change associated with* M_n *and* B_n *the DDS-Brownian motion to* M_n*. The sequence* B_n *converges weakly to a limiting Brownian motion* B*. This holds true in the d-dimensional case, if further* $\langle M_{n,i}, M_{n,j}\rangle_{T^{n,i}_\tau}$ *and* $\langle M_{n,i}, M_{n,j}\rangle_{T^{n,j}_\tau}$ *for all* $i \neq j; i, j \in \{1, ..., d\}$ *converge to zero in probability as* $n \to \infty$*.*

1.1.4 The Cramér-Wold device

For the proofs of weak convergence of weighted sums of d random variables, the following relation to the weak convergence of the d-dimensional random vectors will be useful:

Theorem 1.5 (**Cramér-Wold**)**.** *Let* (X_n) *be a sequence of random vectors in* $\mathbb{R}^d$*. For the weak convergence* $X_n \rightsquigarrow Y$ *with an* $\mathbb{R}^d$*-valued limiting random vector* Y*, a necessary and sufficient condition is given by the weak convergence* $\sum_{k=1}^d \lambda_k X_n^{(k)} \rightsquigarrow \sum_{k=1}^d \lambda_k Y^{(k)}$ *for all* $(\lambda_1, \ldots, \lambda_d) \in \mathbb{R}^d$*.* $X_n^{(k)}$ *and* $Y^{(k)}$ *denote the k-th components of the random vectors* X_n *and* Y*, respectively.*

The theorem is proved in Billingsley [1991] as Theorem 7.7.

1.2 Stable convergence

In this section the notion of stable convergence of sequences of random variables is introduced which will be an essential concept for the development of our limit theory

throughout this work. The concept of stable convergence goes back to Rényi [1963] and results about stable limit theorems were extended in Aldous and Eagleson [1978] and Feigin [1985]. The reason that stable convergence is a key element of our asymptotic theory, is that we will derive results about asymptotic mixed normality. In the case that we have the weak convergence of a sequence of random variables (X_n) to a mixed Gaussian random variable VZ, with $Z \sim \mathbf{N}(\mathbf{0}, \mathbf{1})$, and a strictly positive random variable V, independent of Z, we cannot derive confidence intervals if the distribution of V is unknown. However, if a consistent estimator V_n^2 for the asymptotic variance V^2 is available (in the sense that $V_n^2 \xrightarrow{p} V^2$), the stable convergence will assure that $(X_n, V_n^2) \rightsquigarrow (VZ, V^2)$ jointly and also that $X_n/V_n \rightsquigarrow Z$. Thus, it will be important to prove stable central limit theorems if we are in situations as described above.
In the following, we will present the formal definition and the basic properties of stable convergence of sequences of random variables and give a theorem that will enable us to obtain results about stable convergence in law to mixed Gaussian distributions in the Chapters 3 and 4. The concept of stable convergence also carries over to stochastic processes.

Let (X_n) be a sequence of random variables defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in a Polish space $(E, \mathcal{E})$. We say that the sequence (X_n) converges weakly in $\mathbb{L}^1$ to X if for any bounded random variable Z

$$\lim_{n\to\infty} \mathbb{E}\left[ZX_n\right] = \mathbb{E}\left[ZX\right]$$

holds.

Definition 1.2.1. *For a sub-σ-field $\mathcal{G} \subseteq \mathcal{F}$ the sequence of random variables (X_n) is said to converge $\mathcal{G}$-stably, if there is a random probability measure μ on $(\Omega \times E, \mathcal{G} \otimes \mathcal{E})$ such that*

$$\lim_{n\to\infty} \mathbb{E}\left[Zf(X_n)\right] = \int_{\Omega\times E} \mu(d\omega, dx) Z(\omega) f(x)$$

for all $f \in \mathcal{C}_b(E)$ (continuous and bounded) and $\mathcal{G}$-measurable bounded random variables Z.
If $\mathcal{G} = \mathcal{F}$, we say (X_n) converges stably in law to X ($X_n \overset{st}{\rightsquigarrow} X$).

Remark 1.1. *$\mathcal{G}$-stable convergence is the weak convergence in $\mathbb{L}^1$ of $\mathbb{E}\left[f(X_n)|\mathcal{G}\right]$ for all $f \in \mathcal{C}(E)$ to $\mu \circ f$. This implies convergence in distribution to the probability measure ν defined by $\nu(B) = \int \mu(d\omega, B) \mathbb{1}_{\{X(\omega)\in B\}} \mathbb{P}(d\omega)$. If (X_n) converges stably, the limiting law is $\mu(\Omega, \cdot)$.*

The following proposition states some useful equivalent characterizations of ($\mathcal{G}-$)stable convergence.

Proposition 1.2.2. *(X_n) converges $\mathcal{G}$-stably is equivalent to:*

(i) For every $\mathcal{G}$-measurable random variable Z on Ω, (Z, X_n) converges in law.

(ii) For every $\mathcal{G}$-measurable random variable Z on Ω, (Z, X_n) converges $\mathcal{G}$-stably.

(iii) The sequence (X_n) is tight, and for all $G \in \mathcal{G}$ and $f \in \mathcal{C}(E)$, the sequence $\mathbb{E}\left[\mathbb{1}_G f(X_n)\right]$ converges.

This proposition is proved in Jacod and Shiryaev [2003] as part of Proposition IX.1.4. As indicated in the preliminary motivation, stable convergence can be seen as a stronger version of convergence in law. It is weaker than convergence in probability, but we emphasize that the limit depends on the random variable X itself and not only on the distribution of X.
If (X_n) converges stably to X, X is defined on an extension $(\Omega', \mathcal{F}', \mathbb{P}' = \mu)$ of the original probability space, so that

$$\forall f \in \mathcal{C}(E): \quad \lim_{n\to\infty} \mathbb{E}\left[f(X_n)Z\right] = \mathbb{E}'\left[f(X)Z\right]$$

holds. In the situation described in the introductory paragraph, Gaussian random variables that are independent of $\mathcal{F}$ will appear in the definition of the limiting variable X. In this case we call the extension of the original probability space orthogonal. In particular, if $X_n \overset{st}{\rightsquigarrow} X$ holds with an $\mathcal{F}$-measurable random variable X ($\mathbb{L}^1-$convergence), the foregoing proposition yields that $(X_n, X) \rightsquigarrow (X, X)$ and, hence $(X_n - X) \rightsquigarrow 0$ holds, which implies convergence in probability.
We now give a proposition that shows that stable convergence is the suitable concept to derive feasible central limit theorems and confidence intervals if the asymptotic variances in limit theorems are unknown random, but can be estimated consistently.

Proposition 1.2.3. *Let (X_n, V_n) be real-valued random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$. If $X_n \overset{st}{\rightsquigarrow} X$ with a mixed normal limiting random variable $X \sim \mathbf{N}(0, V^2)$ and $V_n \xrightarrow{p} V$ with V being $\mathcal{F}$-measurable. Then*

$$X_n/V_n \overset{st}{\rightsquigarrow} \mathbf{N}(0,1)$$

holds true.

We use the same denotation expression for mixed normal laws as for normal laws and the difference becomes clear out of the context and by the variances. On the assumptions of the proposition $(X_n, V_n) \overset{st}{\rightsquigarrow} (X, V)$ is implied and the convergence of X_n/V_n follows by the continuous mapping theorem. This proposition is part of Proposition 2.5 in Podolskij and Vetter [2010]. We restricted ourselves to real-valued random variables in the last proposition. A more general version can be found in Jacod and Shiryaev [2003].
For the asymptotic theory of our estimator in the Chapters 3 and 4, the following limit theorem for continuous local martingales will enable us to obtain stable central limit theorems. For the extension of stable convergence to stochastic processes or more precisely to semimartingales the Polish space E in Definition 1.2.1 is given by the Skorokhod space.

Theorem 1.6 (Jacod's theorem: A martingale version). *If $(M_t, \mathcal{F}_t)$ with $0 \le t < \infty$ is a continuous local martingale defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we denote by*

$\mathcal{M}^{\perp}$ the set of bounded $(\mathcal{F}_t)$-adapted martingales orthogonal to $M = (M_t, \mathcal{F}_t)$ what means that $\langle M, M^{\perp}\rangle \equiv 0$. If (X^n) is a sequence of continuous $(\mathcal{F}_t)$-adapted local martingales for which

$$\langle X^n \rangle_t \xrightarrow{p} V_t \quad \forall t \in [0, \infty) \tag{1.2}$$

with a continuous process V holds, the following two conditions

$$\langle X^n, M \rangle_t \xrightarrow{p} 0 \quad \forall t \in [0, \infty) \tag{1.3a}$$

$$\langle X^n, N \rangle_t \xrightarrow{p} 0 \quad \forall t \in [0, \infty) \text{ and } \forall N \in \mathcal{M}^{\perp} \tag{1.3b}$$

are sufficient that (X^n) converges $(\mathcal{F})$-stably in law to W_{V_t}, where W is a standard Brownian motion independent of $\mathcal{F}$.

This theorem is a simplified martingale version of the more general theorem 2–1 in Jacod [1997]. A simplified discrete-time version of that theorem (3–1 in Jacod [1997]) and further motivation and applications of this result can be found in Podolskij and Vetter [2010].
The limiting process in the foregoing Theorem 1.6 is a time-changed Brownian motion. As presented in the last section, every continuous local martingale corresponds to a Dambis, Dubins-Schwarz time-changed Brownian motion. For the sequence (X_t^n) we have a sequence of such Dambis, Dubins-Schwarz Brownian motions $W^n_{\langle X^n \rangle_t}$ that converge weakly to a limiting Brownian motion W_V (cf. Theorem 1.4). The conditions (1.3a) and (1.3b) about the quadratic covariations converging to zero in probability ensure that the weak convergence to W_V is stable.
For one fixed $0 < T < \infty$ we have the result that X_T^n converges stably in law to a centred mixed normal distribution:

$$X_T^n \overset{st}{\rightsquigarrow} \mathbf{N}(0, V_T) \ . \tag{1.4}$$

The independence of the limiting Brownian motion W and (V, Y) for any $\mathcal{F}$-measurable random variable Y assures that (W_{V_T}, Y) has the same law as $(V_T Z, Y)$ with $Z \sim \mathbf{N}(0,1)$ and independent of (V_T, Y).
Note, that in the original theorem 2–1 in Jacod [1997] for semimartingales the same conditions as in our Theorem 1.6 are imposed for the predictable quadratic (co-)variation processes (that coincide with the quadratic (co-)variations for continuous semimartingales). Additionally, a condition that the drift can be neglected asymptotically has to be supposed. The theorem also applies to a multi-dimensional setting that is formulated separately in the next corollary. For this purpose let M^* denote the transpose of a vector M and the $(d \times r)$-dimensional quadratic covariation $\langle M, N^* \rangle_t := \left(\langle M^i, N^j \rangle_t\right)_{ij}$ with $1 \le i \le d$ and $1 \le j \le r$ for a d-dimensional M and r-dimensional N. Recall that convergence in probability of a vector is equivalent to convergence in probability for every component.

Corollary 1.2.4. *Let $(M_t, \mathcal{F}_t)$ be a d-dimensional continuous local martingale and $\mathcal{M}^{\perp}$ again the set of $(\mathcal{F}_t)$-adapted bounded martingales orthogonal to M (to all components).*

A sequence of r-dimensional continuous $(\mathcal{F}_t)$-adapted local martingales (X^n) with

$$\langle X^n, X^{n*}\rangle_t \xrightarrow{p} V_t = \int_0^t w_s w_s^* \, ds \ , \tag{1.5}$$

where w_s is a predictable $\mathbb{R}^r \otimes \mathbb{R}^r$ process, and

$$\langle X^n, M^*\rangle_t \xrightarrow{p} 0 \quad \forall\, t \in [0,\infty) \ , \tag{1.6a}$$

$$\langle X^n, N\rangle_t \xrightarrow{p} 0 \quad \forall\, t \in [0,\infty) \text{ and } \forall\, N \in \mathcal{M}^\perp \ , \tag{1.6b}$$

converges stably in law to the process $\int_0^t w_s dW_s$, where W is a r-dimensional standard Brownian motion independent of $\mathcal{F}$.

Jacod's theorem provides the stable central limit theorem needed for our analysis. Besides Jacod's theorem other stable central limit theorems (e. g. Theorem 3.2 in Hall and Heyde [1980] and van Zanten [2000]) require a certain nesting condition on a sequence of filtrations which is not satisfied here.
The concept of stable convergence will enable us to prove stable weak convergence to mixed Gaussian limits under an equivalent martingale measure $\tilde{\mathbb{P}}$, where the drift processes of the efficient processes are equal to zero identically, and we conclude that the asymptotic law carries over to the case with drift under the original measure $\mathbb{P}$. Stable convergence is commutative with measure change. If we have the result that $Z_n \overset{st}{\rightsquigarrow} m + \mathbf{AVAR} \cdot \mathbf{N}(0,1)$ under $\tilde{\mathbb{P}}$ with a standard Gaussian distribution independent of $\mathcal{F}$, defined on an orthogonal extension of the original probability space and $\mathcal{F}$-measurable bounded random variables m and $\mathbf{AVAR}$, the same convergence holds true under $\mathbb{P}$. Since stable convergence $Z_n \overset{st}{\rightsquigarrow} Z$ implies for all $f \in \mathcal{C}(\mathcal{E})$ and $\mathcal{F}$-measurable bounded random variables X

$$\mathbb{E}\left[Xf(Z_n)\right] = \tilde{\mathbb{E}}\left[(d\mathbb{P}/d\tilde{\mathbb{P}})Xf(Z_n)\right] \to \tilde{\mathbb{E}}'\left[(d\mathbb{P}/d\tilde{\mathbb{P}})Xf(Z)\right] = \mathbb{E}'\left[Xf(Z)\right] \ ,$$

by uniform integrability of $Xf(Z_n)(d\mathbb{P}/d\tilde{\mathbb{P}})$ with $(d\mathbb{P}/d\tilde{P})$ according to (1.1), this result is obtained directly (cf. Mykland and Zhang [2009]).

1.3 Central limit theorems for triangular arrays

1.3.1 Limit theorems for martingale triangular arrays

Definition 1.3.1. *Let $\{S_n, \mathcal{F}_n, n \geq 1\}$ be a centred square-integrable martingale and $Z_n = S_n - S_{n-1}$ $(n \geq 2)$, $(Z_1 = S_1)$ its increments. The discrete predictable compensator of S_n^2*

$$V_n^2 = \sum_{i=1}^n \mathbb{E}[Z_i^2 | \mathcal{F}_{i-1}]$$

is called the conditional variance of the martingale S_n.

We use the notion from Hall and Heyde [1980]. Consider the triangular array $\{S_{n,i}, \mathcal{F}_{n,i}, 1 \leq i \leq k_n\}$ of centred square-integrable martingales with increments $Z_{n,i} = S_{n,i} - S_{n,i-1}$ $(1 \leq i \leq k_n), (S_{n,0} = 0)$. We focus on asymptotics for $k_n \uparrow \infty$ as $n \to \infty$. As in Definition 1.3.1 we denote the conditional variances of $S_{n,i}$ by $V_{n,i}^2 = \sum_{j=1}^{i} \mathbb{E}\left[Z_{n,j}^2 | \mathcal{F}_{n,j-1}\right]$.

Theorem 1.7 (**Martingale central limit theorem**). *Suppose that* η^2 *is an almost surely finite real-valued random variable. On the assumptions that*

$$\max_i |Z_{n,i}| \xrightarrow{p} 0, \tag{1.7}$$

$$\sum_i Z_{n,i}^2 \xrightarrow{p} \eta^2, \tag{1.8}$$

$$\mathbb{E}\left[\max_i Z_{n,i}^2\right] \quad \textit{bounded in } n, \tag{1.9}$$

and

$$\mathcal{F}_{n,i} \subseteq \mathcal{F}_{n+1,i}\ ,\ 1 \leq i \leq k_n,\ n \geq 1\ , \tag{1.10}$$

S_{n,k_n} *converges stably in law to a centred normal distribution with variance* η^2*:*

$$S_{n,k_n} = \sum_i Z_{n,i} \rightsquigarrow \mathbf{N}\left(0, \eta^2\right).$$

Corollary 1.3.2. *The stable convergence in law from Theorem 1.7 also holds true, if instead of the conditions* (1.7) *and* (1.9) *the conditional Lindeberg condition*

$$\forall \lambda > 0: \ \sum_i \mathbb{E}\left[X_{n,i}^2 \mathbb{1}_{[|X_{n,i}|>\lambda]} | \mathcal{F}_{n,i-1}\right] \xrightarrow{p} 0 \tag{C–LB}$$

holds and if condition (1.8) *is replaced by the corresponding assumption on the conditional variance*

$$V_{n,k_n}^2 = \sum \mathbb{E}\left[X_{n,i}^2 | \mathcal{F}_{n,i-1}\right] \xrightarrow{p} \eta^2\ .$$

Corollary 1.3.3. *The conditional Lindeberg condition is implied by the stronger Lyapunov condition*

$$\exists \delta > 0: \ \sum_{i=1}^{n} \mathbb{E}\left[|X_{n,i}|^{2+\delta} \middle| \mathcal{F}_{n,i-1}\right] \xrightarrow{p} 0\ . \tag{C–LY}$$

Proof.

$$\text{If} \quad |Z_{n,i}| = |Z_{n,i} - \mathbb{E}\left[Z_{n,i}\right]| > \epsilon \quad \Rightarrow \quad |Z_{n,i}|^{2+\delta} > |Z_{n,i}|^2 \epsilon^\delta$$

$$\Rightarrow \sum_{i=1}^{n} \mathbb{E}\left[(Z_{n,i})^2 \, \mathbb{1}_{[|Z_{n,i}|>\epsilon]} \middle| \mathcal{F}_{n,i-1}\right] \leq \frac{1}{\epsilon^\delta} \sum_{i=1}^{n} \mathbb{E}\left[|Z_{n,i}|^{2+\delta} \middle| \mathcal{F}_{n,i-1}\right] \; .$$

If the conditional Lyapunov condition (C–LY) holds, the right-hand side converges to zero in probability and hence the left-hand side has to converge to zero in probability, which implies the conditional Lindeberg condition (C–LB). □

The proofs of Theorem 1.7 and Corollary 1.3.2, as well as further information on limit theorems for martingales, can be found in Hall and Heyde [1980] (pages 58 ff.).

Corollary 1.3.4. *Assume that all conditions of Theorem 1.7 except* (1.10) *hold true. If the limit of the (conditional) variances η^2 is measurable in the completions of all $\mathcal{F}_{n,i}$, $1 \leq i \leq k_n$, the non-stable central limit theorem holds true. This particularly includes the case when η^2 is constant.*

This result can be found in Hall and Heyde [1980] in the remarks on page 59 following Corollary 3. 1 and the references cited therein.
Another stable central limit theorem for discrete martingale triangular arrays is obtained as a discrete-time version of Jacod's theorem 1.6:

Corollary 1.3.5. *Assume that $Z_t^n = \sum_{T_{n,i}\leq t} X_{n,i}$ is the endpoint of a discrete martingale and the $X_{n,i}$ are $\mathcal{F}_{T_{n,i}}$-measurable square integrable random variables and $(W_t, \mathcal{F}_t)$ a Brownian motion and $\Delta T_{n,i} = T_{n,i+1} - T_{n,i} \to 0$ as $n \to \infty$. If there exists a predictable process $(v_s)_{s\geq 0}$ such that*

$$\sum_{T_{n,i}\leq t} \mathbb{E}\left[X_{n,i}^2 | \mathcal{F}_{T_{n,i-1}}\right] \xrightarrow{p} \int_0^t v_s^2 \, ds \; , \tag{1.11a}$$

$$\forall \epsilon > 0 : \sum_{T_{n,i}\leq t} \mathbb{E}\left[X_{n,i}^2 \, \mathbb{1}_{\{X_{n,i}>\epsilon\}} | \mathcal{F}_{T_{n,i-1}}\right] \xrightarrow{p} 0 \; , \tag{1.11b}$$

$$\sum_{T_{n,i}\leq t} \mathbb{E}\left[X_{n,i}(W_{T_{n,i}} - W_{T_{n,i-1}}) | \mathcal{F}_{T_{n,i-1}}\right] \xrightarrow{p} 0 \; , \tag{1.11c}$$

$$\sum_{T_{n,i}\leq t} \mathbb{E}\left[X_{n,i}(M_{T_{n,i}} - M_{T_{n,i-1}}) | \mathcal{F}_{T_{n,i-1}}\right] \xrightarrow{p} 0 \; , \tag{1.11d}$$

for all bounded $\mathcal{F}_t$-martingales with $M_0 = 0$ and $\langle W, M\rangle \equiv 0$. Then the following stable convergence of the process Z_t^n holds true:

$$Z_t^n \overset{st}{\rightsquigarrow} Z_t = \int_0^t v_s \, dW_s^\perp \tag{1.12}$$

where $W^\perp$ is a Brownian motion defined on an orthogonal extension of the original probability space.

This Corollary to Jacod's theorem 1.6 is a specific martingale version of Theorem 2. 6 in Podolskij and Vetter [2010] that is itself a simplified version of the more general discrete-time stable limit theorem by Jacod [1997] (theorem 3–1 therein), only that we allow for non-equidistant discrete partitions which does not harm the deduction of Theorem 3–1 from Theorem 2–1 in Jacod [1997]. The conditional Lindeberg-condition (C–LB) is necessary as before in Corollary 1.3.2 and also a convergence condition on the conditional variances. The main difference to the stable limit theorem 1.7 is that the nesting condition on the filtrations (1.10) is replaced by conditions (1.11c) and (1.11d). Usually the reference Brownian motion W is given and "fully generates" the $X_{n,i}$s in the sense that (1.11d) holds.

1.3.2 Limit theorems for triangular arrays under weak dependence

A stochastic process $(Z_j)_{j\in\mathbb{N}_0}$ is said to be stationary, if the random vectors $(Z_{k+1}, Z_{k+2}, \ldots, Z_{k+n})$ and $(Z_0, Z_1, \ldots, Z_n)$ have the same law for all $k \in \mathbb{N}$ and $n \in \mathbb{N}$. For $\sigma-$algebras $\mathcal{A}, \mathcal{B} \subset \mathcal{F}$

$$\alpha(\mathcal{A},\mathcal{B}) = \sup_{A\in\mathcal{A},B\in\mathcal{B}} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\cdot P(B)| \tag{1.13a}$$

$$\phi(\mathcal{A},\mathcal{B}) = \sup_{A\in\mathcal{A},B\in\mathcal{B},\mathbb{P}(A)>0} |\mathbb{P}(B|A) - \mathbb{P}(B)| \tag{1.13b}$$

define two measures of dependence of $\mathcal{A}$ and $\mathcal{B}$. For

$$\mathcal{F}_k = \sigma\left(Z_j,\ 0 \le j \le k\right)$$
$$\mathcal{G}_k = \sigma\left(Z_j, k \le j \le \infty\right)$$

we call

$$\alpha(n) = \sup_{k\in\mathbb{N}} \alpha\left(\mathcal{F}_k, \mathcal{G}_{k+n}\right) \quad \text{and} \quad \phi(n) = \sup_{k\in\mathbb{N}} \phi\left(\mathcal{F}_k, \mathcal{G}_{k+n}\right)$$

the $\alpha-$ and $\phi-$mixing coefficients of the process (Z_j), respectively. For stationary processes

$$\alpha(n) = \alpha\left(\mathcal{F}_0, \mathcal{G}_n\right) \quad \text{and} \quad \phi(n) = \phi\left(\mathcal{F}_0, \mathcal{G}_n\right)$$

holds true. We call (Z_j) strong mixing, if $\alpha(n) \downarrow 0$ as $n \to \infty$. If $\phi(n) \downarrow 0$ as $n \to \infty$ we say the process is $\phi-$mixing and if Z_i and Z_j are independent for $|i - j| > m$ we call it m-dependent. The following implications hold true:

$$\text{m} - \text{dependent} \;\Rightarrow\; \phi - \text{mixing} \;\Rightarrow\; \text{strong mixing} \;\; .$$

For further information on characterizations of non-independent processes and mixing properties we refer to Bradley [2005] and Doukhan [1994]. Strong mixing stationary processes were first considered in 1956 by Rosenblatt as a general class of processes, for which a central limit theorem still holds. The following theorem (Theorem 7.8 from

Durrett [1995]) gives a central limit theorem for stationary strong mixing processes under weak dependence.

Theorem 1.8. *Suppose that (Z_j) is stationary and α-mixing with $\sum_n \alpha_n^{(\delta/(4+2\delta))} < \infty$ and $\mathbb{E}\left[|Z_j|^{2+\delta}\right] < \infty$ for some $\delta > 0$ and that Z_1 has expectation zero. For $S_n = \sum_{j=1}^n Z_j$ the sequence $n^{-1}\mathbb{E}S_n^2$ converges to the limit*

$$\sigma^2 := \mathbb{E}Z_1^2 + 2\sum_{k=1}^{\infty} \mathbb{E}\left[Z_1 Z_k\right] \ , \tag{1.14}$$

where the series converges absolutely. If $\sigma \neq 0$, then

$$\frac{\sigma^{-1}}{\sqrt{n}} S_n \rightsquigarrow \mathbf{N}(0,1) \tag{1.15}$$

holds true.

The analogous result under slightly more restrictive assumptions is given in Theorem 27.4 from Billingsley [1991] and a similar formulation of the theorem is stated in Hall and Heyde [1980] as Corollary 5.1. For the process $Z_n = X_n - X_{n-1}$, where X_n is again stationary and strong mixing, asymptotic normality does not hold which emphasizes that the condition $\sigma \neq 0$ is necessary.
General central limit theorems for triangular arrays under weak dependence is still a vibrant topic of research. Utev [1990] established the following generalization of Lindeberg's central limit theorem:

Theorem 1.9. *Assume $(Z_{n,i}, 1 \leq i \leq k_n)$ is a ϕ-mixing triangular array, such that for every $1 \leq i \leq k_n$, $\mathbb{E}\left[Z_{n,i}\right] = 0$ and $\mathbb{E}\left[Z_{n,i}^2\right] < \infty$. If for every $\varepsilon > 0$ the Lindeberg condition*

$$\frac{1}{\sigma_n^2}\sum_{i=1}^{k_n} \mathbb{E}\left[Z_{n,i}^2 \mathbb{1}_{\{|Z_{n,i}| > \varepsilon\sigma_n\}}\right] \to 0 \ \textit{as } n, k_n \to \infty \tag{LB}$$

holds with $\sigma_n = \mathbb{V}\mathrm{ar}\left(S_n\right)$, the central limit theorem $\sigma_n^{-1} S_n \rightsquigarrow \mathbf{N}(0,1)$ holds.

A related theorem can also be found in Peligrad [1996]. Recall that the Lindeberg condition can always be verified by proving the stronger Lyapunov condition

$$\frac{1}{\sigma_n^{2+\delta}}\sum_{i=1}^{k_n} \mathbb{E}\left[Z_{n,i}^{2+\delta}\right] \to 0 \text{ as } n, k_n \to \infty \ , \tag{LY}$$

for some $\delta > 0$.

1.4 The stochastic Landau symbol

Definition 1.4.1. *Let $(\delta_n)_{n\in\mathbb{N}}$ be a sequence of real values and $(X_n)_{n\in\mathbb{N}}$ a sequence of real-valued random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We write*

$$X_n = \mathcal{O}_p(\delta_n) \quad \textit{(of order } \delta_n \textit{ in probability)} ,$$

if for every $\epsilon > 0$ there is a constant $C_\epsilon \in \mathbb{R}$ and an $n_0 \in \mathbb{N}$ such that

$$\mathbb{P}\left(\left|\frac{X_n}{\delta_n}\right| > C_\epsilon\right) \leq \epsilon \qquad \forall n \geq n_0 \ .$$

If for every $\epsilon > 0$ and for every $\delta > 0$ there is an $n_0 \in \mathbb{N}$ such that

$$\mathbb{P}\left(\left|\frac{X_n}{\delta_n}\right| > \delta\right) \leq \epsilon \qquad \forall n \geq n_0 \ ,$$

we write $X_n = o_p(\delta_n)$ (of smaller order in probability).
$X_n = \mathcal{o}_p(\delta_n)$ thus is equivalent to $X_n/\delta_n \xrightarrow{p} 0$. If we want to highlight the measure $\mathbb{P}$, we write $\mathcal{O}_{\mathbb{P}}(\delta_n)$ and $\mathcal{o}_{\mathbb{P}}(\delta_n)$.

Proposition 1.4.2. *The following arithmetic relations hold similarly to the relations for deterministic Landau symbols: If $X_n = \mathcal{O}_p(\delta_n)$ and $Y_n = \mathcal{O}_p(\gamma_n)$ it holds true that:*

- $X_n \cdot Y_n = \mathcal{O}_p(\delta_n \cdot \gamma_n)$,
- $X_n + Y_n = \mathcal{O}_p\left(\max\left(\delta_n, \gamma_n\right)\right)$,
- $\left|{X_n}^k\right| = \mathcal{O}_p\left({\delta_n}^k\right) \ \forall k \in \mathbb{Q}$.

Analogous arithmetic rules hold for $\mathcal{o}_p(\cdot)$.
If $X_n = \mathcal{O}_p(\delta_n)$ and $Y_n = \mathcal{o}_p(\gamma_n)$, then $X_n \cdot Y_n = \mathcal{o}_P(\delta_n \cdot \gamma_n)$.

Our definition of the stochastic Landau symbol is in line with van der Vaart [1998] (paragraph 2.2), where $X_n = \mathcal{O}_p(1) \Leftrightarrow \lim_{t\to\infty} \sup_{n\in\mathbb{N}} \mathbb{P}\left(|X_n| \geq t\right) = 0$. This means that the sequence of random variables X_n is uniformly bounded and hence uniformly tight. For a sequence of real-valued random variables X_n with $\mathbb{E}X_n = 0$ and $\mathbb{E}X_n^2 = \sigma^2\delta_n^2$ with constant σ^2, Chebyshev's inequality yields

$$\mathbb{P}\left(\left|\frac{X_n}{\delta_n}\right| > C_\epsilon\right) \leq \frac{\sigma^2\delta_n^2}{C_\epsilon^2\delta_n^2} = \frac{\sigma^2}{C_\epsilon^2} \ .$$

Hence, $X_n = O_p(\delta_n)$ holds since for $C_\epsilon = \sqrt{\sigma^2/\epsilon}$ the corresponding condition from Definition 1.4.1 holds true.

1.5 Local asymptotic normality and optimal rates of convergence

A wide variety of statistical models, not only those that incorporate i. i. d. observations, have in common that they behave locally and asymptotically as Gaussian shift models. Hence, for a unified treatment of such models a general concept has been introduced in Le Cam [1960]. One calls a sequence of statistical models or associated statistical experiments locally asymptotically normal (LAN) if their localized likelihood ratio processes converge to that of a normal location model which means that they admit a certain quadratic expansion.
Consider the sequence of statistical experiments $(\mathcal{X}_N, \mathcal{A}_N, \mathbb{P}_\theta^N : \theta \in \Theta)$ with N observations distributed according to the measures $\mathbb{P}_\theta^N$. We give the definition for the special case where θ is a parameter taking values in some open subset $\Theta \subset \mathbb{R}^k$, $k \in \mathbb{N}$, and the underlying probability space $\mathbb{R}^k$ is equipped with the Borel σ-algebra.

Definition 1.5.1 (**LAN**). *A sequence of statistical models $\mathbb{P}_\theta^N$, $\theta \in \Theta$, is locally asymptotically normal (LAN) at θ if there exist matrices r_N and I_θ and random vectors $\Delta_{N,\theta}$ such that $\Delta_{N,\theta} \rightsquigarrow \mathbf{N}(0,1)$ and for every sequence $h_N \to h$*

$$\log\left(\frac{d\mathbb{P}^N_{\theta+r_N^{-1}h_N}}{d\mathbb{P}^N_\theta}\right) = h^\top \Delta_{N,\theta} - \frac{1}{2} h^\top I_\theta h + o_{\mathbb{P}_\theta^N}(1) \ , \tag{1.16}$$

where $h^\top$ denotes the transpose of h.

The first two terms on the right-hand side of (1.16) are the leading terms in the Taylor expansion of the log-likelihood. Important examples for that the LAN property hold are smooth exponential models, autoregressive processes and Gaussian time series models (see e. g. van der Vaart [1998]). In the most well-known examples the norming matrix r_N is diagonal with entries $\sqrt{N}$. From the LAN property one can usually obtain a lower bound for the rate of convergence for any sequence of estimators $\hat{\theta}_N$ of θ thanks to the convolution theorem and the minimax theorem. The convolution theorem was established by Hájek [1970].

Definition 1.5.2. *Let $\theta_N = \theta_0 + r_N^{-1} h_N$. A sequence of estimators $\hat{\theta}_N$ is regular at θ_0 if*

$$\mathcal{L}\left(r_N(\hat{\theta}_N - \theta_N) | \mathbb{P}^N_{\theta_N}\right) \rightsquigarrow Q_{\theta_0}$$

for some limiting law Q_{θ_0} not dependent on (h_N, h).

Theorem 1.10 (**convolution theorem**). *Suppose that the model $\mathbb{P}_\theta^N$ is LAN at θ_0 and $\hat{\theta}_N$ a sequence of estimators regular at θ_0. Then*

$$Q_{\theta_0} = \mathbf{N}\left(0, I_{\theta_0}^{-1}\right) * \nu_{\theta_0} \ ,$$

for some law ν_{θ_0} on $\mathbb{R}^k$. Moreover, $Q_{\theta_0} = \mathbf{N}\left(0, I_{\theta_0}^{-1}\right)$ if and only if $\left(r_N(\hat{\theta}_N - \theta_0)\right) = \Delta_{N,\theta_0} + o_{\mathbb{P}^N_{\theta_0}}(1)$.

The convolution theorem states that the limiting distribution of estimates, if properly normalized, is a convolution of a Gaussian distribution, that depends only on the underlying distributions, and another distribution which depends on the choice of the estimate. Since convolution spreads out the measure $\mathbf{N}\left(0, I_{\theta_0}^{-1}\right)$ a sequence of estimators $\hat{\theta}_N$ is called asymptotically efficient if $Q_{\theta_0} = \mathbf{N}\left(0, I_{\theta_0}^{-1}\right)$. The last element needed to deduce asymptotic lower bounds for the risk is the minimax theorem by Hájek [1972]:

Theorem 1.11 (**local asymptotic minimax theorem**)**.** *Let the sequence of statistical models $\mathbb{P}^N_\theta$ be LAN at θ_0 and $\hat{\theta}_N$ a sequence of estimators (for θ) regular at θ_0. For every symmetric, subconvex and continuous loss function l, it holds true that:*

$$\lim_{c\to 0} \liminf_{N\to\infty} \inf_{\hat{\theta}_N} \sup_{r_N^{-1}|\hat{\theta}_N - \theta_0| \leq c} \mathbb{E}_\theta \left[l\left(r_N \left(\hat{\theta}_N - \theta \right) \right) \right] \geq \mathbb{E}\left[l\left(I_{\theta_0}^{-1/2} Z \right) \right]$$

with a standard Gaussian distributed random vector Z.

Thus, the maximal risk of any estimator in a shrinking neighbourhood of θ_0 is asymptotically bounded below by the Gaussian risk. See van der Vaart [1998], van der Vaart and Wellner [1996], Le Cam [1986] and Le Cam [1972] for further information on LAN and the notion of efficiency of estimators.

2 Dealing with microstructure noise for synchronous observations

2.1 A connatural parametric model: Local asymptotic normality and the optimal rate of convergence

Before introducing nonparametric estimation methods for the quadratic covariation of two Itô processes from discrete observations with additive noise, we consider a connatural parametric model in this section. We aim at estimating the parameter of a constant correlation coefficient of Brownian motions in the statistical model of two synchronously and equidistantly observed Brownian motions. These are observed at sampling times $t_{i,N}$ on the time span $[0,1]$. The index N, emphasizing that we have sequences of sets of observation times, will be omitted in the following for the purpose of a shorter notation. Synchronous observations of the two processes take place at equidistant times with differences $\Delta t_i = t_i - t_{i-1} = \Delta t = 1/N$, $i = 1, \ldots, N$. The observed processes can be written in the following way:

$$\begin{aligned}\tilde{X}_{t_i} &= \int_0^{t_i} dB_t^X + \epsilon_{t_i}^X \ ,\\ \tilde{Y}_{t_i} &= \int_0^{t_i} dB_t^Y + \epsilon_{t_i}^Y \ , \quad i = 0, \ldots, N \ .\end{aligned}$$

We assume the discrete noise processes to be independent of the Brownian motions and independent of each other. We impose an assumption that the noise is i. i. d.-Gaussian:

$$\epsilon_{t_i}^X \overset{iid}{\sim} \mathbf{N}(0, \eta_X^2) \ , \ \epsilon_{t_i}^Y \overset{iid}{\sim} \mathbf{N}(0, \eta_Y^2) \ , i = 0, \ldots, N \ .$$

Constant volatilities σ^X and σ^Y and lengths of the time span T not equal to 1 can be incorporated in the model and in the following analysis, but for a concise notation we keep to the standard Brownian model and state the result for the more general extension at the end of this section.

We will show the local asymptotic normality (LAN) property with rate $N^{-1/4}$ for the correlation coefficient $\rho\, dt = d\langle B^X, B^Y\rangle_t$. With the minimax Theorem 1.11 we conclude that $N^{1/4}$ is a lower bound for the rate of convergence for all estimators of the parameter of interest in this model. The optimal rate also carries over to the more general models of noisy (synchronously or asynchronously) discretely observed Itô processes considered throughout this work. The notion of LAN has been introduced in Section 1.5.

A corresponding result for a constant volatility in the one-dimensional parametric setting

was proved in Gloter and Jacod [2001] with the same rate $N^{-1/4}$. A parametric approach using maximum-likelihood for our two-dimensional model also attains the convergence rate $N^{1/4}$. However, those findings do not necessarily imply that $N^{1/4}$ is a lower bound for the rate of convergence in the bivariate model. There are many examples known where the maximum-likelihood-estimator does not yield rate-optimality (see e. g. Le Cam [1986]) and the statistical model where two Itô processes are observed could be more informative than the one-dimensional model. Conversely, from LAN the asymptotic efficiency of the maximum-likelihood-estimator is obtained under mild regularity conditions.
As a side result, we derive bounds for the asymptotic Fisher information that provide a benchmark for the asymptotic variances of any sequence of estimators for the quadratic covariation, where the dependence of the Fisher information on the correlation coefficient is of particular interest.
We summarize the results of this section in the following Theorem:

Theorem 2.1. *In the model of two synchronously equidistantly observed standard Brownian motions B^X and B^Y with constant correlation ρ and an observation noise described by i. i. d. Gaussian errors with standard deviations η_X and η_Y, the LAN property with $N^{-1/4}$-rate holds ,where N denotes the number of observations in the interval $[0,1]$. Assuming without loss of generality $\eta_X \geq \eta_Y$, we obtain the following lower and upper bound for the asymptotic Fisher information:*

$$\frac{1}{8\eta_X}\left(\frac{1}{(1+\rho)^{3/2}}+\frac{1}{(1-\rho)^{3/2}}\right) \leq I(\rho) \leq \frac{\sqrt{2}}{8}\frac{1}{\sqrt{\eta_X^2+\eta_Y^2}}\left(\frac{1}{(1+\rho)^{3/2}}+\frac{1}{(1-\rho)^{3/2}}\right) . \tag{2.1}$$

In the particular case $\eta_X = \eta_Y = \eta$ the asymptotic Fisher information is given by

$$I(\rho) = \frac{1}{8\eta}\left(\frac{1}{(1+\rho)^{3/2}}+\frac{1}{(1-\rho)^{3/2}}\right) . \tag{2.2}$$

Remark 2.1. *With Theorem 2.1 we prove the LAN property which is one way to conclude the optimal rate of convergence for the estimation problems we are concerned with in this work. It gives, furthermore, bounds for the asymptotic Fisher information that can serve as a benchmark for the asymptotic variances of proposed estimators and in the case of equal noise variances we explicitly obtain the parametric efficiency lower bound for the asymptotic variance.*
The asymptotic Fisher information is enclosed between the 'natural' lower and an intuitive upper bound. We state that the Fisher information (2.2) *has the following asymptotic behaviour in ρ and η:*

$$I(\rho) \to \infty \quad for \quad \rho \to \pm 1 \quad and \quad I(\rho) \to 0 \quad for \quad \eta \to \infty\,.$$

The minimum $\min_\rho I(\rho) = I(0) = (4\eta)^{-1}$ is twice the Fisher information for estimating σ in the univariate case at $\sigma = 1$ (see Gloter and Jacod [2001]). Although the inequalities appearing in the proof for the case of different noise variances are strict, the asymptotic *results do not yield the strict inequalities in* (2.1) *for the lower and upper bound for the asymptotic Fisher information. We suppose that the strict inequalities also hold and a numerical approximation for the Riemann sums using different special values indicated this, too.*

Proof. First we prove the LAN property for the simpler case of equal noise variances $\eta_X = \eta_Y = \eta$ and calculate the asymptotic Fisher information given in formula (2.2). We want to derive the distribution of the increments

$$\Delta\tilde{X}_{t_i} = \int_{t_{i-1}}^{t_i} dB_t^X + \epsilon_{t_i}^X - \epsilon_{t_{i-1}}^X \ , \text{ and } \Delta\tilde{Y}_{t_i} = \int_{t_{i-1}}^{t_i} dB_t^Y + \epsilon_{t_i}^Y - \epsilon_{t_{i-1}}^Y \ ,$$

which follow an MA(1) process here. The constant correlation parameter is denoted by θ in the following. There exists a Brownian motion B independent of $\tilde{X}$ such that the following equation holds:

$$\Delta\tilde{Y}_{t_i} = \int_{t_{i-1}}^{t_i} \theta \, dB_t^X + \sqrt{1-\theta^2} \int_{t_{i-1}}^{t_i} dB_t + \epsilon_{t_i}^Y - \epsilon_{t_{i-1}}^Y \ .$$

Taking this into account we can easily calculate the covariations of the increments:

$$\mathbb{C}\mathrm{ov}(\Delta\tilde{X}_{t_i}, \Delta\tilde{X}_{t_j}) = \mathbb{C}\mathrm{ov}(\Delta\tilde{Y}_{t_i}, \Delta\tilde{Y}_{t_j}) = \begin{cases} \Delta t + 2\eta^2 & \text{if } \quad i = j \\ -\eta^2 & \text{if } |i-j| = 1 \\ 0 & \text{if } |i-j| > 1 \end{cases} ,$$

$$\mathbb{C}\mathrm{ov}(\Delta\tilde{X}_{t_i}, \Delta\tilde{Y}_{t_j}) = \begin{cases} \theta\Delta t & \text{if } \quad i = j \\ 0 & \text{if } \quad i \neq j \end{cases} .$$

The random vector $(\Delta\tilde{X}_{t_1}, \ldots, \Delta\tilde{X}_{t_N}, \Delta\tilde{Y}_{t_1}, \ldots, \Delta\tilde{Y}_{t_N})^t$ has a $(2N \times 2N)$ dimensional covariance matrix

$$\Sigma_\theta = \begin{pmatrix} A_N & D_N \\ D_N & A_N \end{pmatrix}$$

with the $(N \times N)$ matrices

$$A_N = \begin{pmatrix} \Delta t + 2\eta^2 & -\eta^2 & 0 & \ldots & 0 \\ -\eta^2 & \ddots & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & -\eta^2 \\ 0 & \ldots & 0 & -\eta^2 & \Delta t + 2\eta^2 \end{pmatrix}, \ D_N = \begin{pmatrix} \theta\Delta t & 0 & \ldots & \ldots & 0 \\ 0 & \ddots & & & \vdots \\ \vdots & & \ddots & & 0 \\ \vdots & & & \ddots & 0 \\ 0 & \ldots & \ldots & 0 & \theta\Delta t \end{pmatrix}.$$

This special structure of the covariance with the diagonal matrix D_N and the tridiagonal 1-Toeplitz matrix A_N makes it possible to explicitly compute the eigenvalues of Σ_θ. Here

the fact that we assumed the variances of both noise processes to be equal plays an important role.
We write the N-dimensional identity matrix as $\mathbb{1}_N$.
Using a Laplace-expansion, the characteristic polynomials of A_N can be computed by a recursion:

$$\begin{aligned}\det\left(A_N-\lambda\mathbb{1}_N\right) &= \left(\Delta t+2\eta^2-\lambda\right)\det\left(A_{N-1}-\lambda\mathbb{1}_{N-1}\right)+\left(\eta^2\right)^2\det\left(A_{N-2}-\lambda\mathbb{1}_{N-2}\right)\\ &= \sum_{k=0}^{\lfloor\frac{N}{2}\rfloor}(-1)^k\binom{N-k}{k}\left(\Delta t+2\eta^2-\lambda\right)^{N-2k}\left(\eta^2\right)^{2k}\ .\end{aligned}$$

The eigenvalues of A_N are $\lambda_{i,N}=\Delta t+2\eta^2\left(1-\cos\frac{i\pi}{N+1}\right)\ ,\quad i=1,\dots,N$, and because of the simple structure of Σ_θ we can deduce the $2N$ eigenvalues of the covariance matrix directly:

$$\lambda^+_{i,N}(\theta)=\Delta t(1+\theta)+2\eta^2\left(1-\cos\frac{i\pi}{N+1}\right)\ ,\quad i=1,\dots,N\ , \tag{2.3a}$$

$$\lambda^-_{i,N}(\theta)=\Delta t(1-\theta)+2\eta^2\left(1-\cos\frac{i\pi}{N+1}\right)\ ,\quad i=1,\dots,N\ . \tag{2.3b}$$

With the notation

$$\lambda_{j,2N}(\theta)=\begin{cases}\lambda^+_{i,N} & \text{if } j=2i-1\ ,\quad i=1,\dots,N\\ \lambda^-_{i,N} & \text{if } j=2i \qquad\quad ,\quad i=1,\dots,N\end{cases} \tag{2.3c}$$

we can write the $(2N\times 2N)$ diagonal matrix of the eigenvalues as Λ^{2N}_θ with $(\Lambda^{2N}_\theta)_{jj}=\lambda_{j,2N}(\theta)$. Σ_θ can be diagonalized by an $(2N\times 2N)$ orthogonal matrix P^{2N} which is independent of θ. The random vector $P^{2N}\cdot\left(\Delta\tilde{X}_{t_1},\dots,\Delta\tilde{X}_{t_N},\Delta\tilde{Y}_{t_1},\dots,\Delta\tilde{Y}_{t_N}\right)^t$ is centred Gaussian with covariance matrix Λ^{2N}_θ. We define the $2N$-dimensional random vector T^{2N} by

$$\left(T^{2N}\right)_j=\frac{1}{\sqrt{\lambda_{j,2N}(\rho)}}\left(P^{2N}\cdot\left(\Delta\tilde{X}_{t_1},\dots,\Delta\tilde{X}_{t_N},\Delta\tilde{Y}_{t_1},\dots,\Delta\tilde{Y}_{t_N}\right)^t\right)_j\sim\mathbf{N}\left(0\,,\ \frac{\lambda_{j,2N}(\theta)}{\lambda_{j,2N}(\rho)}\right)\ .$$

To prove the LAN property we have to examine the log-likelihood

$$\log\left(\frac{d\mathbb{P}^{2N}_{\rho+N^{-(1/4)}h_N}}{d\mathbb{P}^{2N}_\rho}\right)=-\frac{1}{2}\sum_{j=1}^{2N}\left(\log\left(1+\gamma^{2N}_j\right)-\left(T^{2N}\right)^2_j\frac{\gamma^{2N}_j}{\gamma^{2N}_j+1}\right)$$

where

$$\gamma^{2N}_j=\frac{\lambda_{j,2N}\left(\rho+N^{-1/4}h_N\right)}{\lambda_{j,2N}(\rho)}-1=\frac{\Delta t\cdot N^{-1/4}h_N}{\lambda_{j,2N}(\rho)}\ .$$

The proof is now analogous to the one-dimensional case in Gloter and Jacod [2001] and using Theorem VIII-3.32 in Jacod and Shiryaev [2003] it remains to show that

$$\sup_{1\le j\le 2N}|\gamma_j^{2N}|\to 0 \quad \text{and} \quad \sum_{j=1}^{2N}\left(\gamma_j^{2N}\right)^2\to 2h^2 I(\rho)\ . \tag{2.4}$$

The first condition is obviously fulfilled. To prove the second one, we write the sum of the squares as a Riemann sum:

$$\sum_{j=1}^{2N}\left(\gamma_j^{2N}\right)^2=\frac{N^{1/2}h_N^2\,(\Delta t)^2}{\left(\eta^2\right)^2\pi}\left(S_N+\tilde{S}_N\right)$$

with

$$S_N=\frac{\pi}{N}\sum_{j=1}^{N}\frac{1}{\left(2\left(1-\cos\frac{j\pi}{N+1}\right)+\frac{\Delta t(1+\rho)}{\eta^2}\right)^2},\ \tilde{S}_N=\frac{\pi}{N}\sum_{j=1}^{N}\frac{1}{\left(2\left(1-\cos\frac{j\pi}{N+1}\right)+\frac{\Delta t(1-\rho)}{\eta^2}\right)^2}\ .$$

Using the lower and upper Darboux sums of the corresponding integrals

$$J=\int_0^{\pi}\frac{1}{\left(2\left(1-\cos z\right)+\frac{\Delta t(1+\rho)}{\eta^2}\right)^2}\,dz\ ,\ \tilde{J}=\int_0^{\pi}\frac{1}{\left(2\left(1-\cos z\right)+\frac{\Delta t(1-\rho)}{\eta^2}\right)^2}\,dz\ ,$$

this yields the following inequality for the Riemann sums S_N:

$$J\le S_N\le J+\frac{\pi}{N}\frac{1}{N\left(4+\frac{\Delta t(1+\rho)}{\eta^2}\right)^2}-\frac{\pi}{N}\frac{1}{\left(2\left(1-\cos\frac{\pi}{N+1}\right)+\frac{\Delta t+(1+\rho)}{\eta^2}\right)^2}\ ,$$

and the analogous one for the Riemann sums $\tilde{S}_N$ with the integral $\tilde{J}_N$. The integrals can be computed explicitly and

$$\frac{N^{-3/2}h_N^2}{\left(\eta^2\right)^2\pi}\left(J+\tilde{J}\right)=\frac{N^{-3/2}h_N^2}{\left(\eta^2\right)^2\pi}\left[\frac{\pi\left(2+\frac{\Delta t(1+\rho)}{\eta^2}\right)}{\left(\frac{\Delta t(1+\rho)}{\eta^2}\left(\frac{\Delta t(1+\rho)}{\eta^2}+4\right)\right)^{3/2}}+\frac{\pi\left(2+\frac{\Delta t(1-\rho)}{\eta^2}\right)}{\left(\frac{\Delta t(1-\rho)}{\eta^2}\left(\frac{\Delta t(1-\rho)}{\eta^2}+4\right)\right)^{3/2}}\right]$$

holds true. Since $h_N\to h$, we can deduce from the preceding inequalities for both addends the convergence

$$\sum_{j=1}^{2N}\left(\gamma_j^{2N}\right)^2\to\frac{h^2}{4\eta}\left(\frac{1}{(1+\rho)^{3/2}}+\frac{1}{(1-\rho)^{3/2}}\right)=2h^2I(\rho) \tag{2.5}$$

with the Fisher information

$$I(\rho)=\frac{1}{8\eta}\left(\frac{1}{(1+\rho)^{3/2}}+\frac{1}{(1-\rho)^{3/2}}\right)\ . \tag{2.6}$$

We continue the proof with the generalization for different noise variances. If the noise variances are not equal, $\eta_X^2 \neq \eta_Y^2$, the covariance matrix can be written as

$$\Sigma_\theta = \begin{pmatrix} A_N & D_N \\ D_N & B_N \end{pmatrix}$$

with the same diagonal matrix D_N as before and two tridiagonal 1-Toeplitz matrices A_N and B_N with the same structure as before where A_N has the entries $\Delta t + 2\eta_X^2$ on the main diagonal and correspondingly, B_N the entries $\Delta t + 2\eta_Y^2$. The eigenvalues of A_N and B_N have been deduced before and are denoted by $\lambda_X^{(i,N)}$ and $\lambda_Y^{(i,N)}$ here, which emphasizes the dependence on η_X and η_Y, respectively. Because of the special structure of A_N and B_N, that are in particular symmetric and commutative, they share the same eigenvectors. We can use this to calculate the $2N$ eigenvalues of Σ_θ, denoted by $\xi_+^{(i)}, \xi_-^{(i)}, i = 1, \ldots, N$:

$$\xi_+^{(i)} = \frac{\lambda_X^{(i)} + \lambda_Y^{(i)}}{2} + \sqrt{\left(\frac{\lambda_X^{(i)} - \lambda_Y^{(i)}}{2}\right)^2 + \theta^2 (\Delta t)^2}\ ,$$

$$\xi_-^{(i)} = \frac{\lambda_X^{(i)} + \lambda_Y^{(i)}}{2} - \sqrt{\left(\frac{\lambda_X^{(i)} - \lambda_Y^{(i)}}{2}\right)^2 + \theta^2 (\Delta t)^2}\ .$$

We have dropped the index N of the eigenvalues here.

Lemma 2.1.1. *If we assume $\eta_X > \eta_Y$, the following inequalities hold:*

$$\frac{\lambda_X^{(i)} + \lambda_Y^{(i)}}{2} + \theta\Delta t < \xi_+^{(i)} < \lambda_X^{(i)} + \theta\Delta t\ , \tag{2.7a}$$

$$\lambda_X^{(i)} - \theta\Delta t < \xi_-^{(i)} < \frac{\lambda_X^{(i)} + \lambda_Y^{(i)}}{2} - \theta\Delta t\ . \tag{2.7b}$$

Proof. If $\eta_X > \eta_Y$ for the eigenvalues $\lambda_X^{(i)} > \lambda_Y^{(i)}$ holds for all $i \in \{1, \ldots, N\}$. Thus

$$\xi_+^{(i)} < \frac{\lambda_X^{(i)} + \lambda_Y^{(i)}}{2} + \sqrt{\left(\frac{\lambda_X^{(i)} - \lambda_Y^{(i)}}{2}\right)^2 + \left(\lambda_X^{(i)} - \lambda_Y^{(i)}\right)\theta\Delta t + \theta^2 (\Delta t)^2} = \lambda_X^{(i)} + \theta\Delta t$$

holds and analogously the lower bound for $\xi_-^{(i)}$ is obtained by adding the mixed term to the expression under the square root. The other bounds are obvious. $\square$

In the following, we define

$$\gamma_+^{(i)} = \frac{\xi_+^{(i)}\left(\rho + N^{-1/4} h_N\right)}{\xi_+^{(i)}(\rho)} - 1 > 0 \quad \text{and} \quad \gamma_-^{(i)} = \frac{\xi_-^{(i)}\left(\rho + N^{-1/4} h_N\right)}{\xi_-^{(i)}(\rho)} - 1 < 0$$

in analogy to the case of equal noise variances. We use the preceding lemma to obtain bounds for these coefficients and show the LAN property with the same rate $N^{-1/4}$ as above, including bounds for the Fisher information.

Proposition 2.1.2. *If $\eta_X > \eta_Y$ the following inequalities hold:*

$$\frac{N^{-\frac{1}{4}}h_N\Delta t + \frac{\lambda_Y^{(i)}-\lambda_X^{(i)}}{2}}{\lambda_X^{(i)} + \rho\Delta t} < \gamma_+^{(i)} < \frac{N^{-\frac{1}{4}}h_N\Delta t}{\frac{\lambda_X^{(i)}+\lambda_Y^{(i)}}{2} + \rho\Delta t} \tag{2.8a}$$

and

$$\frac{-N^{-\frac{1}{4}}h_N\Delta t}{\frac{\lambda_X^{(i)}+\lambda_Y^{(i)}}{2} - \rho\Delta t} < \gamma_-^{(i)} < \frac{-N^{-\frac{1}{4}}h_N\Delta t + \frac{\lambda_Y^{(i)}-\lambda_X^{(i)}}{2}}{\lambda_X^{(i)} - \rho\Delta t}\,. \tag{2.8b}$$

Proof. Using the inequality (2.7a) in the preceding Lemma 2.1.1 we obtain the lower bound for $\gamma_+^{(i)}$. We can deduce the upper bound using again the right-hand side of inequality (2.7a) in the last inequality. The bounds for $\gamma_-^{(i)}$ follow analogously. □

Now we are able to prove the LAN property in the same way as for the case of equal noise variances using the preceding inequalities. Because of Proposition 2.1.2, the inequalities

$$\sum_{i=1}^N \left(\gamma_+^{(i)}\right)^2 + \sum_{i=1}^N \left(\gamma_-^{(i)}\right)^2 < \sum_{i=1}^N \left(\frac{N^{-\frac{1}{2}}h_N^2\,(\Delta t)^2}{\left(\frac{\lambda_X^{(i)}+\lambda_Y^{(i)}}{2} + \rho\Delta t\right)^2} + \frac{N^{-\frac{1}{2}}h_n^2\,(\Delta t)^2}{\left(\frac{\lambda_X^{(i)}+\lambda_Y^{(i)}}{2} - \rho\Delta t\right)^2} \right)$$

and

$$\begin{aligned} &\sum_{i=1}^N \left(\gamma_+^{(i)}\right)^2 + \sum_{i=1}^N \left(\gamma_-^{(i)}\right)^2 \\ &> \sum_{i=1}^N \left(\frac{N^{-\frac{1}{2}}h_N^2\,(\Delta t)^2 + \left(\frac{\lambda_Y^{(i)}-\lambda_X^{(i)}}{2}\right)^2}{\left(\lambda_X^{(i)} + \rho\Delta t\right)^2} + \frac{N^{-\frac{1}{2}}h_N^2\,(\Delta t)^2 + \left(\frac{\lambda_Y^{(i)}-\lambda_X^{(i)}}{2}\right)^2}{\left(\lambda_X^{(i)} - \rho\Delta t\right)^2} \right) \\ &> \sum_{i=1}^N \left(\frac{N^{-\frac{1}{2}}h_N^2\,(\Delta t)^2}{\left(\lambda_X^{(i)} + \rho\Delta t\right)^2} + \frac{N^{-\frac{1}{2}}h_N^2\,(\Delta t)^2}{\left(\lambda_X^{(i)} - \rho\Delta t\right)^2} \right) \end{aligned}$$

hold. In the lower bound the cross terms drop out.
Using those inequalities, the proof reduces to the method used before for the equal noise variance case where we found that (Riemann) sums of this type can be approximated by integrals. We just have to do this calculation twice for the upper and the lower bound

changing only the constants in the denominator of the integrated function and obtain the convergence to $2h^2\underline{I(\rho)}$ and $2h^2\overline{I(\rho)}$, respectively, with the lower and upper bound for $I(\rho)$ stated in formula (2.1). □

The structure of the eigenvalues (2.3a) and (2.3b) of the covariance matrix already indicate and give an heuristic why the usual $\sqrt{N}$-rate cannot be attained for this estimation problem. The first addend with the parameter of interest θ is of order $\Delta t = N^{-1}$ and dominates the second addend for the first $\sqrt{N}$ eigenvalues. For all other eigenvalues the second addend, that is due to the observation noise, becomes dominating what can be seen using a Taylor expansion for the cosine.
In the model with general T, σ^X, σ^Y the entries of the matrices A_N on the diagonal become $T(\sigma^X)^2/N + 2\eta_X^2$ and $T(\sigma^Y)^2/N + 2\eta_Y^2$ for B_N. The entries of D_N have an additional factor $\sigma^X\sigma^Y$. LAN can be proven with small adaptations and

$$\frac{1}{8\eta_X}\left(\frac{1}{(T(\max(\sigma_X^2, \sigma_Y^2) + \rho\sigma^X\sigma^Y))^{3/2}} + \frac{1}{(T(\max(\sigma_X^2, \sigma_Y^2) - \rho\sigma^X\sigma^Y))^{3/2}}\right),$$

$$\frac{\sqrt{2}}{8}\frac{1}{\sqrt{\eta_X^2 + \eta_Y^2}}\left(\frac{1}{(T(\min(\sigma_X^2, \sigma_Y^2) + \rho\sigma^X\sigma^Y))^{3/2}} + \frac{1}{(T(\min(\sigma_X^2, \sigma_Y^2) - \rho\sigma^X\sigma^Y))^{3/2}}\right)$$

constitute the lower and upper bound for the Fisher information.
We have proven the LAN property with rate $N^{-1/4}$ in this simplified model and thus we will be able to conclude that our generalized multiscale estimator that will be derived in Chapter 4 is rate-optimal. Even in the less informative statistical model of synchronous equidistant observations the minimax theorem ensures that $N^{1/4}$ is a lower bound for the convergence rates of any sequence of estimators.

2.2 Subsampling estimators for the integrated covariance

From now on we consider more general continuous semimartingales than in the parametric setting of the last section.

Assumption 1 (efficient processes). *On a filtered probability space* $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$*, the so-called efficient processes* $X = (X_t)_{t\in\mathbb{R}^+}$ *and* $Y = (Y_t)_{t\in\mathbb{R}^+}$ *are Itô processes defined by the following stochastic differential equations:*

$$dX_t = \mu_t^X\,dt + \sigma_t^X\,dB_t^X\ ,$$
$$dY_t = \mu_t^Y\,dt + \sigma_t^Y\,dB_t^Y\ ,$$

with two $(\mathcal{F}_t)$*–adapted standard Brownian motions* B^X *and* B^Y *and* $\rho_t\,dt = d\langle B^X, B^Y\rangle_t$*. The drift processes* μ_t^X *and* μ_t^Y *are* $(\mathcal{F}_t)$*–adapted locally bounded stochastic processes and the spot volatilities* σ_t^X *and* σ_t^Y *and* ρ_t *are assumed to be* $(\mathcal{F}_t)$*–adapted with continuous* *paths. We assume strictly positive volatilities and the Novikov condition* $\mathbb{E}\left[\exp\left((1/2)\int_0^T(\mu^{\cdot}/\sigma^{\cdot})_t^2\,dt\right)\right] < \infty$ *for* X *and* Y*.*

Note that adapted continuous stochastic processes are always locally bounded (cf. page 140 in Revuz and Yor [1991]) and since we aim at estimating the quadratic covariation $\langle X, Y\rangle_T$ over a fixed time span $[0,T]$, we can use continuity and boundedness of the paths for our further analysis. The Novikov condition allows us to remove the drift by a measure change with the Girsanov Theorem 1.2 for in other respects very general volatility and drift processes.
In this section we are concerned with synchronous discrete observations of X and Y with an additive microstructure noise what is precised in the following assumption.

Assumption 2.1 (**observations with noise**)**.** *The processes* X *and* Y *are both observed at times* $t_j^{(n)}, j = 0,\ldots,n$ *on* $[0,T]$ *with additive discrete noise processes:*

$$\tilde{X}_{t_j^{(n)}} = X_{t_j^{(n)}} + \epsilon^X_{t_j^{(n)}}\ ,\ \tilde{Y}_{t_j^{(n)}} = Y_{t_j^{(n)}} + \epsilon^Y_{t_j^{(n)}}\quad j = 0,\ldots,n\ .$$

The microstructure noise processes are assumed to be i. i. d. , mutually independent and also independent of the efficient processes. Furthermore, the errors are centred, noise variances denoted by η_X^2 *and* η_Y^2 *and fourth moments are finite. For the observation scheme the supremum of instants* $\delta_n := \sup_{1\le j\le n}(t_j^{(n)} - t_{j-1}^{(n)}, t_0^{(n)}, T - t_n^{(n)})$*, also called mesh size, tends to zero as we consider asymptotics for* $n\to\infty$*. It can tend to zero slower than the average time instant* T/n*, but not to slow. We assume there exists a constant* $0 < \alpha \le 1/9$*, such that* $\delta_n = \mathcal{O}\left(n^{-8/9-\alpha}\right)$ *holds true.*

In the following, we drop the superscript (n) of observation times $t_j, j = 0,\ldots,n$. For non-noisy observations and a mesh size $\delta_n \to 0$ as $n\to\infty$ the quadratic covariation $\langle X,Y\rangle_T = \int_0^T \rho_t\sigma_t^X\sigma_t^Y\,dt$ can be estimated with the realized covariance $\sum_{j=1}^n(X_{t_j} - X_{t_{j-1}})(Y_{t_j} - Y_{t_{j-1}})$ for which the stable central limit theorem

$$\sqrt{n}\left(\sum_{j=1}^n\left(X_{t_j} - X_{t_{j-1}}\right)\left(Y_{t_j} - Y_{t_{j-1}}\right) - \langle X,Y\rangle_T\right) \overset{st}{\rightsquigarrow} \mathbf{N}\left(0, T\int_0^T(\rho_t^2+1)(\sigma_t^X\sigma_t^Y)^2\,dG(t)\right)$$

holds which is covered by the more general result of Theorem 3.1 in Chapter 3. The limit of the quadratic variations of observation times G is defined below in Proposition 2.2.1. In the model imposed by Assumption 1 and Assumption 2.1, the realized covariance and also the realized volatilities do not provide consistent estimators for the quadratic (co-)variations any more. The variance due to noise

$$\mathbb{V}\mathrm{ar}_{X,Y}\left(\sum_{j=1}^n\left(\tilde{X}_{t_j} - \tilde{X}_{t_{j-1}}\right)\left(\tilde{Y}_{t_j} - \tilde{Y}_{t_{j-1}}\right)\right) = 4n\,\eta_X^2\eta_Y^2\ ,$$

conditional on the paths of the efficient processes denoted by $\mathbb{V}\mathrm{ar}_{X,Y}$, increases linearly with the number of observations n. For the univariate realized volatilities $\sum_{j=1}^n(\tilde{X}_{t_j} -$

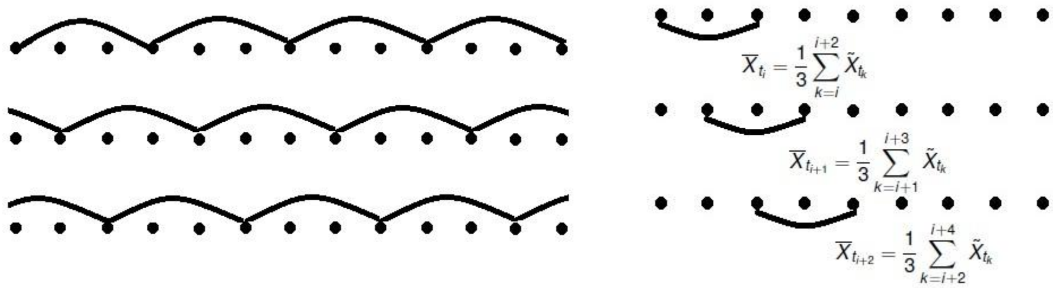


Figure 2.1: Sketch of the subsampling approach.

$\tilde{X}_{t_{j-1}})^2 = 2n\eta_X^2 + \mathcal{O}_p\left(\sqrt{n}\right)$ holds (see Zhang et al. [2005]). Thus, the bias of the realized volatilities and the variances due to noise can be reduced when they are evaluated using less data than available at the highest frequency what has been called sparse-sampling in Zhang et al. [2005]. This aligns with the common former practice for high-frequency financial data before the methods presented in this and the next section had been suggested. From the signature plot in Figure 0.1 one would simply have taken a realized volatility at some lower frequency as an estimator for the quadratic variation of the underlying efficient process. In the additive i. i. d. noise model this frequency can be chosen optimally in the sense that the mean square error is minimized, but this value depends on the unknown noise variances and the integrated quarticity $\int_0^T \left(\sigma_t^X\right)^4 dG(t)$ in the univariate and $\int_0^T (\rho_t^2+1)(\sigma_t^X\sigma_t^Y)^2\, dG(t)$ in the bivariate case, respectively.
At this point, we start an overview on the existing methods with an intuitive estimator, which will be called subsampling estimator in this work, which has been proposed for the univariate estimation of integrated volatility as the "second best approach" in Zhang et al. [2005]. We extend the approach to the bivariate setting what is for the synchronous case very similar to the univariate. It can be motivated from two points of view that are both sketched in Figure 2.1. On the left-hand side we have visualized that one can calculate simultaneously lower frequent realized covariances on subsamples, e. g. to the lag three in Figure 2.1, and (post-)average them to obtain the final subsampling estimator.

$$\widehat{\langle X,Y\rangle}_T^{sub} = \frac{1}{i}\sum_{j=i}^{n}\left(\tilde{X}_{t_j}-\tilde{X}_{t_{j-i}}\right)\left(\tilde{Y}_{t_j}-\tilde{Y}_{t_{j-i}}\right)\;. \tag{2.9a}$$

This motivation is in line with the former practice and proposes to use an average instead of one single lower frequency realized measure and has been given in Zhang et al. [2005]. The same estimator arises as the usual realized covariance calculated from the time series on that we have run a linear filter before what means that non-noisy observations at a time t_j are estimated with a pre-average of noisy observations at times $t_j,\ldots,t_{j+i}$ for some i. This is sketched on the right-hand side in Figure 2.1 for $i=3$. Passing over to increments leads to telescoping sums and we end up finally with the same subsampling

estimator.
Since on the Assumption 2.1 there is no bias due to noise, the bivariate estimator already corresponds to the “first best approach” from Zhang et al. [2005] whereas in the univariate case a bias-correction completes the two scales realized volatility (TSRV):

$$\widehat{\langle X \rangle}_T^{TSRV} = \frac{1}{i} \sum_{j=i}^{n} \left(\tilde{X}_{t_j} - \tilde{X}_{t_{j-i}} \right)^2 - \frac{1}{2n} \sum_{i=1}^{n} \left(\tilde{X}_{t_j} - \tilde{X}_{t_{j-1}} \right)^2 . \tag{2.9b}$$

The subsample frequency i will be chosen dependent on n but we leave out indices in the following to guarantee a comprehensible notation. The variance of the subsampling estimator can be written

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left(\widehat{\langle X, Y \rangle}_T^{sub} \right) &= \frac{1}{i^2} \sum_{j=i}^{n} \sum_{k=i}^{n} \mathbb{C}\mathrm{ov}\Big(\left(\tilde{X}_{t_j} - \tilde{X}_{t_{j-i}} \right) \left(\tilde{Y}_{t_j} - \tilde{Y}_{t_{j-i}} \right) , \\
&\qquad\qquad \left(\tilde{X}_{t_k} - \tilde{X}_{t_{k-i}} \right) \left(\tilde{Y}_{t_k} - \tilde{Y}_{t_{k-i}} \right) \Big) \\
&= \frac{1}{i^2} \sum_{j=i}^{n} \sum_{k=i}^{n} \Big(\mathbb{C}\mathrm{ov}\left(ⓔ_j, ⓔ_k \right) + \mathbb{C}\mathrm{ov}\left(ⓜ_j, ⓜ_k \right) + \mathbb{C}\mathrm{ov}\left(ⓥ_j, ⓥ_k \right) + \mathbb{C}\mathrm{ov}\left(ⓝ_j, ⓝ_k \right) \Big) \\
&= \mathbb{V}\mathrm{ar}_{\mathrm{n}} + \mathbb{V}\mathrm{ar}_{\mathrm{cross}} + \mathbb{V}\mathrm{ar}_{\mathrm{dis}}
\end{aligned} \tag{2.10}$$

with the four uncorrelated terms

$$\begin{aligned}
ⓔ_j &= \left(X_{t_j} - X_{t_{j-i}} \right) \left(Y_{t_j} - Y_{t_{j-i}} \right) , \\
ⓜ_j &= \left(X_{t_j} - X_{t_{j-i}} \right) \left(\epsilon_{t_j}^{Y} - \epsilon_{t_{j-i}}^{Y} \right) , \\
ⓥ_j &= \left(\epsilon_{t_j}^{X} - \epsilon_{t_{j-i}}^{X} \right) \left(Y_{t_j} - Y_{t_{j-i}} \right) , \\
ⓝ_j &= \left(\epsilon_{t_i}^{X} - \epsilon_{t_{j-i}}^{X} \right) \left(\epsilon_{t_j}^{Y} - \epsilon_{t_{j-i}}^{Y} \right) .
\end{aligned}$$

There is a trade-off between the discretization variance $\mathbb{V}\mathrm{ar}_{\mathrm{dis}}$ that is of order i/n and the variance due to noise $\mathbb{V}\mathrm{ar}_{\mathrm{n}}$ being of order n/i^2. The variance due to cross terms $\mathbb{V}\mathrm{ar}_{\mathrm{cross}}$ converges to zero in probability as $i \to \infty$, $n \to \infty$, $i/n \to 0$. Thus, choosing $i = c_{sub} n^{2/3}$ with a constant c_{sub} the mean square error is minimized and of order $n^{-1/3}$. The subsampling estimator (2.9a) is a consistent, asymptotically unbiased estimator. The rate of convergence $n^{1/6}$, however, is slow and does not attain the optimal rate $n^{1/4}$ determined in the last section. This can be remedied with the methods that have further refined the subsampling estimator. In the following, we focus on a multiscale approach on which the methods developed in Chapter 4 of this work are based on. The multiscale realized covariance (MSRC), and the univariate multiscale realized volatility (MSRV) introduced in Zhang [2006], are weighted sums of subsampling estimators with

M_n different subsampling frequencies $i = 1, \ldots, M_n$:

$$\widehat{\langle X,Y\rangle}_T^{multi} = \sum_{i=1}^{M_N} \frac{\alpha_{i,M_N}^{opt}}{i} \sum_{j=i}^{n} \left(\tilde{X}_{t_j} - \tilde{X}_{t_{j-i}}\right)\left(\tilde{Y}_{t_j} - \tilde{Y}_{t_{j-i}}\right) \,, \tag{2.11a}$$

$$\widehat{\langle X\rangle}_T^{multi} = \sum_{i=1}^{M_N} \frac{\alpha_{i,M_N}^{opt}}{i} \sum_{j=i}^{n} \left(\tilde{X}_{t_j} - \tilde{X}_{t_{j-i}}\right)^2 \,. \tag{2.11b}$$

The two alternative methods that have been considered a lot in that strand of literature are presented to the reader in the next section. The weights of the multiscale estimators are chosen such that the estimator is asymptotically unbiased and the error due to noise minimized. These discrete weights given later in (4.14) are the same for the bivariate case and the univariate (cf. Zhang [2006]). We will abstain from giving a more general class of weights, determined by weight functions on a grid, what is provided in Zhang [2006] and for the two other methods in the literature.
The mean square error of the multiscale realized covariance (2.11a) can be split again in three uncorrelated parts that stem from discretization, microstructure noise and cross terms and end-effects. They are of order M_n/n, n/M_n^3, and M_n^{-1}, respectively. Hence, a choice $M_n = c_{multi}\sqrt{n}$ leads to a rate-optimal $n^{1/4}$-consistent estimator.
The following stable central limit theorems for the multiscale realized covariance (2.11a) and the subsampling estimator (2.9a) are implied by the general main result of this work in Theorem 4.1 and Corollary 4.2.2 in Chapter 4.

Proposition 2.2.1. *On Assumptions 1 and 2.1 and that* $(n/T)\sum_i (t_i^{(n)} - t_{i-1}^{(n)})^2$ *converges to a continuously differentiable limiting function* G *and the difference quotients converge uniformly to* G' *on* $[0,T]$*, the multiscale realized covariance* (2.11a) *and the subsampling estimator* (2.9a) *converge stably in law to centred mixed normal limiting random variables:*

$$n^{1/4}\left(\widehat{\langle X,Y\rangle}_T^{multi} - \langle X,Y\rangle_T\right) \overset{st}{\rightsquigarrow} \mathbf{N}\left(0\,,\, \mathbf{AVAR}_{multi,syn}\right) \,, \tag{2.12a}$$

$$n^{1/6}\left(\widehat{\langle X,Y\rangle}_T^{sub} - \langle X,Y\rangle_T\right) \overset{st}{\rightsquigarrow} \mathbf{N}\left(0\,,\, \mathbf{AVAR}_{sub,syn}\right) \,, \tag{2.12b}$$

with

$$\begin{aligned}\mathbf{AVAR}_{multi,syn} &= c_{multi}^{-3} 24\eta_X^2\eta_Y^2 + c_{multi}\frac{26}{35}T\int_0^T G'(t)(\rho_t^2+1)(\sigma_t^X\sigma_t^Y)^2\,dt \\ &\quad + c_{multi}^{-1}\frac{12}{5}\left(\eta_X^2\eta_Y^2 + \eta_X^2\int_0^T(\sigma_t^Y)^2\,dt + \eta_Y^2\int_0^T(\sigma_t^X)^2\,dt\right) \,,\end{aligned} \tag{2.12c}$$

$$\mathbf{AVAR}_{sub,syn} = c_{sub}^{-2} 4\eta_X^2\eta_Y^2 + c_{sub}\frac{2}{3}T\int_0^T G'(t)(\rho_t^2+1)(\sigma_t^X\sigma_t^Y)^2\,dt\ . \tag{2.12d}$$

2.3 Alternative estimation methods for the quadratic covariation

This section is devoted to two alternative approaches to deal with microstructure noise in a latent semimartingale model for estimating quadratic (co-)variations. We give a short overview on the methods, existing results and the associated references. The methods have been developed for integrated volatility estimation but there has been some extensions to a multivariate setting and as for the subsampling method the estimators for synchronous observations are straight extended versions of the univariate ones. We also refer to the first extensions that allow for application to asynchronous observation schemes that have been presented in concurrent literature during the development of this work.

Realized kernels or autocovariance estimators

The second approach proposed to estimate the quadratic variation in an additive noise model with a latent efficient continuous semimartingale that followed the one by Zhang et al. [2005] has been called kernel-based estimation by the authors Barndorff-Nielsen et al. [2008a]. Since this method is often referred to as autocovariance-based estimation by other authors we have stated both names. The reason for the latter is that the estimators arise as linear combinations of autocovariances whereas the subsampling estimator is a linear combination of realized covariances. We will keep to the notation of the last section. A wide class of kernel functions can be plugged in the general estimator that is given as

$$K(\tilde{X},\tilde{Y}) = \sum_{h=0}^{H_n}\mathfrak{K}\left(\frac{h}{H_n+1}\right)\Big(\sum_{j=h+1}^{n}\left(\tilde{X}_{t_j}-\tilde{X}_{t_{j-1}}\right)\left(\tilde{Y}_{t_{j-h}}-\tilde{Y}_{t_{j-h-1}}\right) + \mathbb{1}_{h\neq 0}\left(\tilde{X}_{t_{j-h}}-\tilde{X}_{t_{j-h-1}}\right)\left(\tilde{Y}_{t_j}-\tilde{Y}_{t_{j-1}}\right)\Big)$$

with a kernel $\mathfrak{K}$ for $T=1$ on $[0,1]$, that satisfies the following conditions:

- $\mathfrak{K}$ is twice continuously differentiable with $\int_0^1\mathfrak{K}^2(x)\,dx<\infty$, $\int_0^1(\mathfrak{K}'(x))^2\,dx<\infty$, $\int_0^1(\mathfrak{K}''(x))^2\,dx<\infty$.
- $\mathfrak{K}(0)=1$, $\mathfrak{K}'(0)=0$.
- $\mathfrak{K}(1)=0$ and $\mathfrak{K}'(0)=\mathfrak{K}'(1)=0$.

In the univariate case Barndorff-Nielsen et al. [2008a] have proved that for $H=c\cdot\sqrt{n}$ these estimators attain the optimal convergence rate $n^{1/4}$ and are asymptotically distributed

according to a mixed Gaussian law. We conjecture without proof that analogous properties carry over to the bivariate synchronous setting for the above given estimators and kernels. The last of the conditions on the kernel function restricts our choice to so-called flat-top kernels. When substituting this condition by the condition that $\int_{-\infty}^{\infty} \mathfrak{K}(x) \exp(i\lambda x) dx \geq 0$ $\forall \lambda \in \mathbb{R}$ holds, and choosing $H = c \cdot n^{3/5}$, one ends up with an estimator that does not achieve the optimal rate $n^{1/4}$, but a slower $n^{1/5}$-rate of convergence. However, the estimator guarantees positive semi-definiteness of the resulting estimates for the covariance matrix in a multivariate setting what is an important feature from an applied point of view in a multivariate model. This has been shown in Barndorff-Nielsen et al. [2008b].
The asymptotic properties of the autocovariance estimators are deduced by separating the terms due to noise contamination, the latent semimartingale and cross terms:

$$K(\tilde{X}, \tilde{Y}) = \underbrace{K(X, Y)}_{\xrightarrow{p} \langle X, Y \rangle_1} \underbrace{+ K(X, \epsilon^Y) + K(\epsilon^X, Y) + K(\epsilon^X, \epsilon^Y)}_{\xrightarrow{p} 0} \; .$$

For appropriate choices of H and $\mathfrak{K}$, the first addend will converge to the quadratic covariation and the others to zero in probability. On regularity assumptions similar to the ones above in Proposition 2.2.1, and even slightly more general, for $H = c \cdot n^{3/5}$ and an aforementioned chosen kernel, it holds true that:

$$n^{1/5} \left(K(\tilde{X}, \tilde{Y}) - \int_0^1 \rho_s \sigma_s^X \sigma_s^Y \, ds \right) \xrightarrow{\mathcal{L}_{st}} \mathbf{N} \left(\frac{|\mathfrak{K}''(0)|}{c^2} \Omega_{XY} \, , \, \mathbf{AVAR} \right)$$

where

$$\mathbf{AVAR} = 2c \left(\int_0^1 \mathfrak{K}^2(x) \, dx \right) \int_0^1 g^2 \left((1 + \rho_s^2)(\sigma_s^X \sigma_s^Y)^2 \, ds \right) \circ f du$$

and the estimated covariance matrices are positive semi-definite. When H_n is chosen such that $\sqrt{n}/H_n \to 0$, the error due to noise is completely 'smoothed out' as can be seen in the asymptotic variance. The transformation with g and f in the integral is redundant in an equidistant case.
Since this approach is compatible with a previous-tick interpolation to refresh times, that are defined and explained as 'closest synchronous approximation' (3.4) in Chapter 3 of this work, the above result stays valid in a non-synchronous setting. This combined estimation method adopted from Barndorff-Nielsen et al. [2008b] has set the current standard for integrated covariance estimation from non-synchronous noisy high-frequency observations.
For this reason, we will compare the methodology developed in this thesis later on to this previous-tick interpolation refresh time and kernel-based estimation method.

A pre-average approach

The pre-average method, that has been proposed for integrated volatility estimation in this kind of statistical models by Podolskij and Vetter [2009], and has been further examined and generalized in Jacod et al. [2009], relies on the idea of pre-averaging the noisy observations and calculate a realized volatility from the averaged values. It is accomplished differently than the simple linear filter described in the motivation of the subsampling estimator above. In particular, the noisy observations are at first averaged on fixed blocks not leading to telescoping sums. Now, consider equidistant observations at times $t_j^{(n)} = j/n, j = 0, \ldots, n$ and the notation introduced in the last section. The final bivariate estimator is attained with the "modulated realised covariance"

$$\mathbf{MRC} = \frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\sum_{i=0}^{n-k_n+1}\bar{Y}_i^n \bar{X}_i^n \;,$$

with

$$\bar{Y}_i^n = \left(\sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right)\left(\tilde{Y}_{\frac{i+j}{n}} - \tilde{Y}_{\frac{i+j-1}{n}}\right)\right) \;,$$

$$\bar{X}_i^n = \left(\sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right)\left(\tilde{X}_{\frac{i+j}{n}} - \tilde{X}_{\frac{i+j-1}{n}}\right)\right) \;,$$

that has to be bias-corrected similarly as the TSRV

$$\widehat{\langle X,Y\rangle}_T^{pre-avg} = \mathbf{MRC} - \frac{\psi_1}{\theta^2\psi_2}\frac{1}{2n}\sum_{i=0}^{n-1}\left(\tilde{X}_{\frac{i+1}{n}} - \tilde{X}_{\frac{i}{n}}\right)\left(\tilde{Y}_{\frac{i+1}{n}} - \tilde{Y}_{\frac{i}{n}}\right) \;.$$

This estimator also provides a rate-optimal estimator and feasible stable central limit theorems have been established for the univariate version and the equidistant synchronous bivariate case. Different weight functions g can be inserted that fulfill the following conditions:

- g is continuous and piecewise continuously differentiable.
- g' is Lipschitz-continuous.
- $g(0) = g(1) = 0$.
- $\int_0^1 g^2(s)ds > 0$.

The block-lengths are chosen $k_n = \theta\sqrt{n}$ with a constant θ and

$$\psi_2 := \int_0^1 g^2(s)\,ds \quad , \quad \psi_1 := \int_0^1 \left(g'(s)\right)^2\,ds \;.$$

The original pre-average estimator has incorporated local averages of the form

$$\begin{aligned}\bar{X}_i^n &= \frac{1}{k_n}\left(\sum_{j=k_n/2}^{k_n-1}\tilde{X}_{\frac{i+j}{n}} - \sum_{j=0}^{k_n/2-1}\tilde{X}_{\frac{i+j}{n}}\right)\\ &= \frac{1}{k_n}\sum_{l=0}^{k_n/2-1}\left(\tilde{X}_{\frac{i+l+k_n/2}{n}} - \tilde{X}_{\frac{i+l}{n}}\right)\\ &= \frac{1}{k_n}\sum_{l=0}^{k_n/2-1}\left(\sum_{r=0}^{l+k_n/2-1}\left(\tilde{X}_{\frac{i+r+1}{n}} - \tilde{X}_{\frac{i+r}{n}}\right)\right)\\ &= \frac{1}{k_n}\sum_{j=1}^{k_n-1}\left(\tilde{X}_{\frac{i+j}{n}} - \tilde{X}_{\frac{i+j-1}{n}}\right)\min\left(j, k_n - j\right)\ ,\end{aligned}$$

where k_n is assumed to be odd without loss of generality and is therefore contained in the above generalized version with the weight function $g(x) = \min\left(x, 1-x\right)$. In this case $\psi_1 = 1,\ \psi_2 = 1/12$ holds true. The pre-averaged values can be written as the sum of pre-averages of the latent efficient process and noise corruption: $\bar{X}_i^n = \bar{X}_i^{n,eff} + \overline{\epsilon^X}_i^n$. Those two addends are of order $\sqrt{k_n/n}$ and $\sqrt{1/k_n}$ in probability, respectively. Here, we already see that a choice $k_n = \theta\sqrt{n}$ balances the two sources of errors.
In an equidistant setting for $T = 1$ the following consistency and asymptotic mixed normality results hold true:

$$\widehat{\langle X, Y\rangle}_1^{pre-avg} \xrightarrow{p} \int_0^1 \rho_s\sigma_s^X\sigma_s^Y\,ds\ ,$$

$$n^{1/4}\left(\mathbf{MRC} - \int_0^1 \rho_s\sigma_s^X\sigma_s^Y\,ds\right) \xrightarrow{\mathcal{L}_{st}} \mathbf{N}\left(0\,,\,\mathbf{AVAR}\right)\ ,$$

with the asymptotic variance

$$\begin{aligned}\mathbf{AVAR} = \frac{2}{\psi_2^2}\Big(&\frac{151}{80640}\theta\int_0^1(1+\rho_s^2)(\sigma_s^X\sigma_s^Y)^2\,ds\\ &+ \frac{1}{96\theta}\int_0^1\left((\sigma_s^X\eta_Y)^2 + (\sigma_s^Y\eta_X)^2 + 2\eta_{XY}\rho_s\sigma_s^X\sigma_s^Y\right)ds + \frac{1}{6\theta^3}\left(\eta_X^2\eta_Y^2 + \eta_X^2\eta_Y^2\eta_{XY}\right)\Big)\ ,\end{aligned}$$

where η_{XY} is the covariance $\mathbb{E}\left[\epsilon_{j/n}^X\epsilon_{j/n}^Y\right]$.
These results are taken from Christensen et al. [2010]. In this article several refinements of the estimator are given. Choosing k_n larger such that $\sqrt{n}/k_n \to 0$, estimated covariance matrices are positive semi-definite but the convergence rate decreases. Furthermore, the authors provide a first combination of the pre-average approach with the Hayashi-Yoshida estimator, that is considered in Chapter 3 of this work, and prove that this method can attain the optimal rate in the non-synchronous and noisy setting.

All three approaches have achieved estimators that are rate-optimal, asymptotically unbiased and asymptotically mixed Gaussian distributed for the integrated volatility what carries over to a synchronous bivariate setting. Asymptotic variances will look similar as well, only the constants that stem also from the used weight functions can differ depending on the exact forms of the estimators and the inserted weights. Apart from that, the methods mainly differ in treatment of end-effects. The estimators are in fact very closely related and for specific weight functions or kernels they can be transformed to each other. Especially the transition between the kernel estimator from Barndorff-Nielsen et al. [2008a] to the general multiscale estimator in Zhang [2006] is obtained directly by using the second derivative of the kernel as weight function f or the multiscale estimator. Ignoring end-effects, the original pre-average estimator corresponds to an autocovariance estimator using the Parzen kernel $\mathfrak{K}(x) = 1 - 6x^2 + 6x^3\mathbb{1}_{[0,1/2]}(x) + 2(1-x)^3\mathbb{1}_{[1/2,1]}(x)$ and the multiscale estimator with the weights (4.14) on that our combined method will be grounded to the cubic kernel $\mathfrak{K}(x) = 1 - 3x^2 + 2x^3$ (see Christensen et al. [2010] and Barndorff-Nielsen et al. [2008a]). In principle, combined methods as developed in Chapter 4 for the most general noisy case can be based on any of the three approaches.

3 Dealing with asynchronous sampling schemes for non-noisy observations

3.1 The Hayashi-Yoshida estimator and a related synchronization algorithm

In this chapter we consider the estimation of the integrated covariance $\langle X, Y\rangle_T$ of two Itô processes X and Y as defined in Assumption 1 from discrete observations following non-synchronous sampling schemes, but without additional microstructure noise.
We impose the following regularity assumptions on the underlying asynchronous sampling schemes:

Assumption 2 (**observations**). *The deterministic observation times $\mathcal{T}^{X,n} = \{0 \leq t_0^{(n)} < t_1^{(n)} < \ldots < t_n^{(n)} \leq T\}$ of X and $\mathcal{T}^{Y,m} = \{0 \leq \tau_0^{(m)} < \tau_1^{(m)} < \ldots < \tau_m^{(m)} \leq T\}$ of Y are assumed to be regular in the following sense:*

(a) There exists a constant $0 < \alpha \leq 1/3$ such that

$$\delta_n^X = \sup_{i\in\{1,\ldots,n\}} \left(\left(t_i^{(n)} - t_{i-1}^{(n)}\right), t_0^{(n)}, T - t_n^{(n)}\right) \quad = \mathcal{O}\left(n^{-2/3-\alpha}\right)\ , \tag{3.1a}$$

$$\delta_m^Y = \sup_{j\in\{1,\ldots,m\}} \left(\left(\tau_j^{(m)} - \tau_{j-1}^{(m)}\right), \tau_0^{(m)}, T - \tau_m^{(m)}\right) = \mathcal{O}\left(m^{-2/3-\alpha}\right)\ . \tag{3.1b}$$

(b) There exists a constant $0 < \alpha \leq 1/9$ such that

$$\delta_n^X = \sup_{i\in\{1,\ldots,n\}} \left(\left(t_i^{(n)} - t_{i-1}^{(n)}\right), t_0^{(n)}, T - t_n^{(n)}\right) \quad = \mathcal{O}\left(n^{-8/9-\alpha}\right)\ , \tag{3.1c}$$

$$\delta_m^Y = \sup_{j\in\{1,\ldots,m\}} \left(\left(\tau_j^{(m)} - \tau_{j-1}^{(m)}\right), \tau_0^{(m)}, T - \tau_m^{(m)}\right) = \mathcal{O}\left(m^{-8/9-\alpha}\right)\ . \tag{3.1d}$$

We consider asymptotics where the number of observations of X and Y are assumed to be of the same asymptotic order $n = \mathcal{O}(m)$ and $m = \mathcal{O}(n)$ and express that shortly by $n \sim m$.

For synchronous data $n = m$ and $t_i^{(n)} = \tau_i^{(n)}$ for all $i \in \{0, \ldots, n\}$ holds. In the non-synchronous case the number of observations $(n+1)$ of X and $(m+1)$ of Y differ and the sets of observation times $\mathcal{T}^{X,n}$ also contain times $t_i^{(n)} \notin \mathcal{T}^{Y,m}$ and $\tau_j^{(n)} \notin \mathcal{T}^{X,n}$. We work within the general model where synchronous observation times can take place and hence $\mathcal{T}^{Y,m}$ and $\mathcal{T}^{X,n}$ are not assumed to be disjoint. In the following, we omit the

upper indices (n) and (m) for the observation times and the times defined below that depend on sequences of sampling schemes.
Although the sequences of observation times are modeled deterministically, we remark that the case of random sampling times that are independent of the observed processes is included in that analysis regarding the conditional law given the observation times.
We use the short notation $\Delta X_{t_i}, i = 1, \ldots, n$ from now on for increments $X_{t_i} - X_{t_{i-1}}$ and analogously for Y. Hayashi and Yoshida [2005] have proved the consistency of their estimator

$$\widehat{\langle X, Y \rangle}_T^{(HY)} = \sum_{i=1}^{n} \sum_{j=1}^{m} \Delta X_{t_i} \Delta Y_{\tau_j} \mathbb{1}_{[\min(t_i,\tau_j) > \max(t_{i-1},\tau_{j-1})]} \,,$$

where the product terms include all increments of the processes with overlapping observation time intervals, for a similar model of discretely observed Itô diffusions with deterministic correlation, drift and volatility functions. Consistency directly carries over to our setting including random correlation, drift and volatility processes. The estimator is also in our setting, furthermore, unbiased if drift terms are zero and else asymptotically unbiased. Hayashi and Yoshida [2008] have further proven that on stronger regularity assumptions on the design their estimator is asymptotically Gaussian distributed.
For a combination of the strategy of the Hayashi-Yoshida estimator with techniques to handle noise contamination in Chapter 4, we focus on an alternative useful method to deal with the asynchronicity of the data. It was introduced in Palandri [2006] (which he calls pseudo-aggregation). This method provides an iterative algorithm to rewrite the estimator of Hayashi and Yoshida without indicator functions. This can be done by aggregation of addends for which partial sums are telescoping. A first simple rewriting of the Hayashi-Yoshida estimator is obtained by taking the sum of the products of all increments of X with the telescoping sums of aggregated observed increments of Y for that observation time instants overlap with the according observation time instant of X (or in the symmetric way):

$$\begin{aligned}
\widehat{\langle X, Y \rangle}_T^{(HY)} &= \sum_{i=1}^{n} \Delta X_{t_i} \left(\sum_{j \in \{1,\ldots,m\}} \Delta Y_{\tau_j} \mathbb{1}_{[\min(t_i,\tau_j) > \max(t_{i-1},\tau_{j-1})]} \right) \\
&= \sum_{j=1}^{m} \Delta Y_{\tau_j} \left(\sum_{i \in \{1,\ldots,n\}} \Delta X_{t_i} \mathbb{1}_{[\min(t_i,\tau_j) > \max(t_{i-1},\tau_{j-1})]} \right) \,.
\end{aligned}$$

Defining the next-tick interpolation $t_{i,+} := \min_{0 \le j \le m} (\tau_j | \tau_j \ge t_i)$ and the previous-tick interpolation $t_{i,-} := \max_{0 \le j \le m} (\tau_j | \tau_j \le t_i)$, the last expression can be illustrated

$$\widehat{\langle X, Y \rangle}_T^{(HY)} = \sum_{i=1}^{n} \Delta X_{t_i} \left(Y_{t_{i,+}} - Y_{t_{i-1,-}} \right) \,.$$

The algorithm we will use is a more enhanced method to aggregate the data in an adequate way. For this purpose $(N+1)$ sets $\mathcal{H}^i$ and $\mathcal{G}^i$ are constructed, where $N < \min(n, m)$, each set including one or more than one observation time of X and Y, respectively. This method to construct a joint grid for the observations of the two processes is described by the iterative Algorithm 3.1.

The Algorithm 3.1 stops after $(N+1)$ steps when the last observation time is reached.

first step:

- for $t_0 < \tau_0$ and $\mu_0 = \min(w \in \{1,\dots,n\} | \tau_0 \le t_w)$:

$$\mathcal{H}^0 = \{t_0, \dots, t_{\mu_0}\} \quad \text{and} \quad \mathcal{G}^0 = \{\tau_0\}$$

$$q_1 = \begin{cases} \mu_0 + 1 & \text{if } \tau_0 = t_{\mu_0} \\ \mu_0 & \text{if } \tau_0 < t_{\mu_0} \end{cases} \qquad \text{and} \quad r_1 = 1$$

- for $t_0 = \tau_0$:

$$\mathcal{H}^0 = \{t_0\} \quad \text{and} \quad \mathcal{G}^0 = \{\tau_0\}$$

$$q_1 = 1 \quad \text{and} \quad r_1 = 1$$

- for $t_0 > \tau_0$ and $w_0 = \min(l \in \{1,\dots,m\} | t_0 \le \tau_l)$:

$$\mathcal{H}^0 = \{t_0\} \quad \text{and} \quad \mathcal{G}^0 = \{\tau_0, \dots, \tau_{w_0}\}$$

$$q_1 = 1 \quad \text{and} \quad r_1 = \begin{cases} w_0 + 1 & \text{if } t_0 = \tau_{w_0} \\ w_0 & \text{if } t_0 < \tau_{w_0} \end{cases}$$

ith step (given $\mathcal{H}^{i-1}$ and $\mathcal{G}^{i-1}$):

- for $t_{q_i} < \tau_{r_i}$ and $\mu_i = \min(w \in \{q_i + 1,\dots,n\} | \tau_{r_i} \le t_w)$:

$$\mathcal{H}^i = \{t_{q_i}, \dots, t_{\mu_i}\} \quad \text{and} \quad \mathcal{G}^i = \{\tau_{r_i}\}$$

$$q_i \dashrightarrow \begin{cases} q_{i+1} = \mu_i + 1 & \text{if } \tau_{r_i} = t_{\mu_i} \\ q_{i+1} = \mu_i & \text{if } \tau_{r_i} < t_{\mu_i} \end{cases} \qquad \text{and} \quad r_i \dashrightarrow r_{i+1} = r_i + 1$$

- for $t_{q_i} = \tau_{r_i}$:

$$\mathcal{H}^i = \{t_{q_i}\} \quad \text{and} \quad \mathcal{G}^i = \{\tau_{r_i}\}$$

$$q_i \dashrightarrow q_{i+1} = q_i + 1 \quad \text{and} \quad r_i \dashrightarrow r_{i+1} = r_i + 1$$

- for $t_{q_i} > \tau_{r_i}$ and $w_i = \min(l \in \{r_i + 1,\dots,m\} | t_{q_i} \le \tau_l)$:

$$\mathcal{H}^i = \{t_{q_i}\} \quad \text{and} \quad \mathcal{G}^i = \{\tau_{r_i}, \dots, \tau_{w_i}\}$$

$$q_i \dashrightarrow q_{i+1} = q_i + 1 \quad \text{and} \quad r_i \dashrightarrow \begin{cases} r_{i+1} = w_i + 1 & \text{if } t_{q_i} = \tau_{w_i} \\ r_{i+1} = w_i & \text{if } t_{q_i} < \tau_{w_i} \end{cases}$$

Algorithm 3.1: Iterative algorithm for construction of the joint grid from asynchronous data.

We pass over from the original observations to the sums of observed increments $X^{\mathcal{H}^i}$ over sets $\mathcal{H}^i$ and $Y^{\mathcal{G}^i}$ over sets $\mathcal{G}^i$, respectively. The observations are grouped together so that the resulting realized covariance estimator

$$\sum_{i=0}^{N} X^{\mathcal{H}^i} Y^{\mathcal{G}^i} = \sum_{i=1}^{n}\sum_{j=1}^{m} \Delta X_{t_i} \Delta Y_{\tau_j} \mathbb{1}_{[\min(t_i,\tau_j) > \max(t_{i-1},\tau_{j-1})]}$$

calculated from the 'synchronized' observations

$$X^{\mathcal{H}^i} = \sum_{t_j \in \mathcal{H}^i} \Delta X_{t_j} \ , \quad Y^{\mathcal{G}^i} = \sum_{\tau_j \in \mathcal{G}^i} \Delta Y_{\tau_j} \ , \quad i \in \{0, \ldots, N\} \ .$$

for the integrated covariance will coincide with the one by Hayashi and Yoshida [2005]. We use a different illustration of this estimator compared to Palandri [2006] making use of telescoping sums.
With the denotation expressions from Algorithm 3.1

$$\begin{aligned} \mu_i &= \max(k|t_k \in \mathcal{H}^i), & w_i &= \max(k|\tau_k \in \mathcal{G}^i) \ \text{ and} \\ q_i &= \min(k|t_k \in \mathcal{H}^i), & r_i &= \min(k|\tau_k \in \mathcal{G}^i) \ \ , i \in \{0, \ldots, N\} \end{aligned}$$

and for the purpose of a simpler notation

$$\begin{aligned} X_{g_i} &= X_{t_{\mu_i}}, & Y_{\gamma_i} &= Y_{\tau_{w_i}} & &, i \in \{0, \ldots, N\} \text{ and} \\ X_{l_i} &= X_{t_{q_i - 1}}, & Y_{\lambda_i} &= Y_{\tau_{r_i - 1}} & &, i \in \{1, \ldots, N\} \end{aligned}$$

with $l_0 := t_0$, $\lambda_0 := \tau_0$, $X^{\mathcal{H}^i}$ and $Y^{\mathcal{G}^i}$ can be written as telescoping sums $X^{\mathcal{H}^i} = (X_{g_i} - X_{l_i})$, $Y^{\mathcal{G}^i} = (Y_{\gamma_i} - Y_{\lambda_i})$. This leads to

$$\widehat{\langle X, Y \rangle}_T^{(HY)} = \sum_{i=1}^{N} (X_{g_i} - X_{l_i})(Y_{\gamma_i} - Y_{\lambda_i}) \ , \tag{3.2}$$

where summation starts with $i = 0$ or $i = 1$ since the addend for $i = 0$ is always zero. Although we use this specific new illustration throughout this chapter, we will call this realized covariance of our synchronized observations Hayashi-Yoshida estimator in the following. In this notation g_i denotes the greatest and l_i the last observation time before the least element of the set $\mathcal{H}^i$ and analogously γ_i and λ_i of $\mathcal{G}^i$.

Example
An illustration of the application of Algorithm 3.1 to observations is given in Figure 3.2. In this example, we have $\mathcal{H}^0 = \{t_0\}, \mathcal{G}^0 = \{\tau_0\}, \mathcal{H}^1 = \{t_1, t_2, t_3\}, \mathcal{G}^1 = \{\tau_1\}, \mathcal{H}^2 = \{t_3\}, \mathcal{G}^2 = \{\tau_2, \tau_3\}, \mathcal{H}^3 = \{t_4, t_5, t_6\}, \mathcal{G}^3 = \{\tau_4\}, \mathcal{H}^4 = \{t_6, t_7\}, \mathcal{G}^4 = \{\tau_5\}, \mathcal{H}^5 = \{t_7, t_8\}, \mathcal{G}^5 = \{\tau_6\}, \mathcal{H}^6 = \{t_8\}, \mathcal{G}^6 = \{\tau_7, \tau_8\}, \mathcal{H}^7 = \{t_9\}, \mathcal{G}^7 = \{\tau_8, \tau_9\}, \mathcal{H}^8 = \{t_{10}\}, \mathcal{G}^8 = \{\tau_9, \tau_{10}\}$.

The example emphasizes the important fact that the sets $\mathcal{H}^i$ and $\mathcal{G}^i$ are in general not disjoint and the maxima of consecutive sets can be the same time points. The minimum of a successive set can as well equal the maximum of the prevenient. For further examples

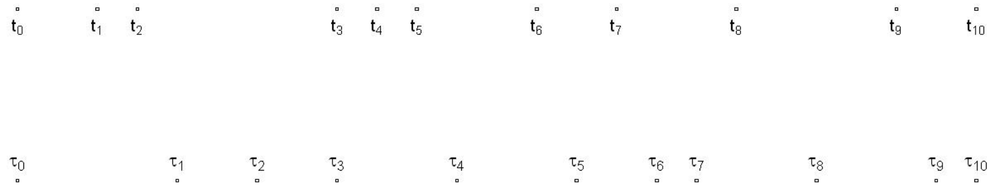


Figure 3.2: Example for synchronization using Algorithm 3.1.

see Palandri [2006]. Of course the example is just for illustration and the number of observations is restricted and much smaller than in practice. The synchronization of $n+1=11$ and $m+1=11$ observations leads to $N+1=9$ synchronized observations in this example.

The 'translation' of the Hayashi-Yoshida approach with that iterative algorithm will be useful for our analysis of noise terms in Chapter 4. In particular, this construction will enable us to deal with the noise contamination by applying subsampling techniques. The fact that we obtain $(N+1) < \min(n,m)+1$ synchronized observations indicates heuristically that the efficiency of such techniques of covariance estimation depends on the number of observations available for the process observed at a lower frequency. By Assumption 2(a) we restrict us to the case that n and m are of the same order. Thus for the suprema of times between two observations

$$\delta_n^X = \mathcal{O}\left(N^{-2/3-\alpha}\right) \quad \text{and} \quad \delta_n^Y = \mathcal{O}\left(N^{-2/3-\alpha}\right)$$

holds.
In the next sections it will be shown that on Assumption 1 and 2(a) the estimator (3.2) is $\sqrt{N}$-consistent and, on further assumptions on the asymptotic behavior of the asynchronous sampling schemes, asymptotically Gaussian distributed. This rate-optimal estimator in the absence of noise is an adequate starting point for the development of a combined method in the most general asynchronous and noisy setting in Chapter 4. Using standard interpolation methods such an estimator cannot be obtained.
Another recent approach to deal with non-synchronous discrete observations has been proposed by Barndorff-Nielsen et al. [2008b]. This method is also related to our approach. The so-called refresh times are the cumulative sums of waiting times until both processes are observed. Assume that in the ith step of Algorithm 3.1 $t_{q_i} < \tau_{r_i}$ holds. Then the next observation times of X are grouped together ending with the first observation time $t_{\mu_i-1} < \tau_{r_i} \leq t_{\mu_i}$ greater or equal than τ_{r_i}. Then we start the next comparison step and compare this last observation time grouped to the set $\mathcal{H}^i$ to τ_{r_i+1}, except for the case where two synchronous observations appeared, where we compare the two

following times. Since in the completely asynchronous case at the refresh times only one of the two processes is observed, the refresh time method used in Barndorff-Nielsen et al. [2008b] includes a previous-tick interpolation for the unobserved process at the refresh times. Refresh times provide the 'closest synchronous approximation' to the asynchronous sampling schemes that we define in Proposition 3.4 below. The number of refresh times which are denoted in this work by $T_i, i = 0, \ldots, N$, equals the number of sets constructed by pseudo-aggregation. The previous-tick interpolation, however, causes a negative bias due to asynchronicity when calculating the simple realized covariance estimator based on the refresh time and previous-tick approach and it does not equal the estimator of Hayashi-Yoshida. The reason for this bias is that, due to the previous-tick interpolation, products of increments with overlapping observation time instants fall out of the realized covariance. The pseudo-aggregation Algorithm 3.1 used in this work corresponds to the refresh time method when replacing the previous-tick interpolation by a next-tick interpolation for the right end points of refresh time instants. Then, the resulting realized covariance of 'synchronized observations'

$$\begin{aligned}\widehat{\langle X, Y\rangle}_T^{(HY)} &= \sum_{i=1}^{N} \left(X_{g_i} - X_{l_i}\right)\left(Y_{\gamma_i} - Y_{\lambda_i}\right) \\ &= \sum_{i=1}^{N} \left(X_{T_{i,+}} - X_{T_{i-1,-}}\right)\left(Y_{T_{i,+}} - Y_{T_{i-1,-}}\right) \end{aligned} \tag{3.3}$$

coincides with the Hayashi-Yoshida estimator and has no bias due to asynchronicity.

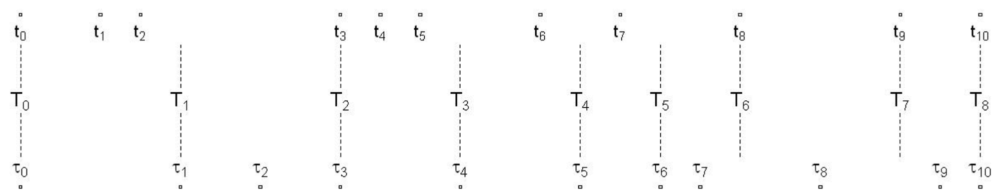


Figure 3.3: Example for synchronization using Algorithm 3.1 including refresh times.

Figure 3.3 visualizes refresh times T_i, $i = 0, \ldots, 8$ for our above given example. For this example the realized covariance calculated with refresh times previous-tick interpolated values equals

$$\begin{aligned}&(X_{t_2} - X_{t_0})(Y_{\tau_1} - Y_{\tau_0}) + (X_{t_3} - X_{t_2})(Y_{\tau_3} - Y_{\tau_1}) + (X_{t_5} - X_{t_3})(Y_{\tau_4} - Y_{\tau_3}) + \\ &(X_{t_6} - X_{t_5})(Y_{\tau_5} - Y_{\tau_4}) + (X_{t_7} - X_{t_6})(Y_{\tau_6} - Y_{\tau_5}) + (X_{t_8} - X_{t_7})(Y_{\tau_7} - Y_{\tau_6}) + \\ &\qquad (X_{t_9} - X_{t_8})(Y_{\tau_8} - Y_{\tau_7}) + (X_{t_{10}} - X_{t_9})(Y_{\tau_{10}} - Y_{\tau_8})\end{aligned}$$

and is biased downwards, whereas (3.2) yields

$$\begin{aligned}&(X_{t_3}-X_{t_0})(Y_{\tau_1}-Y_{\tau_0})+(X_{t_3}-X_{t_2})(Y_{\tau_3}-Y_{\tau_1})+(X_{t_6}-X_{t_3})(Y_{\tau_4}-Y_{\tau_3})+\\&(X_{t_7}-X_{t_5})(Y_{\tau_5}-Y_{\tau_4})+(X_{t_8}-X_{t_6})(Y_{\tau_6}-Y_{\tau_5})+(X_{t_8}-X_{t_7})(Y_{\tau_8}-Y_{\tau_6})+\\&\qquad(X_{t_9}-X_{t_8})(Y_{\tau_9}-Y_{\tau_7})+(X_{t_{10}}-X_{t_9})(Y_{\tau_{10}}-Y_{\tau_8})\ ,\end{aligned}$$

which is an unbiased estimator for observations of processes according to Assumption 1, when drift terms are assumed to be zero.

3.2 Asymptotics of the estimator: A stable central limit theorem

In this section the basic elements for an analysis of the asymptotic properties of the estimator (3.2) are developed. We focus on a result on the asymptotic distribution of the estimator.

Proposition 3.2.1. *If we define $T_i := \min(g_i,\gamma_i)$, $i=0,\dots,N$, the set $\mathcal{T}^{syn} = \{T_0,\dots,T_N\}$ induces a partition of the time span $[0,T]$ in the sense that $\dot{\bigcup}_i[T_i,T_{i+1}) = [T_0,T-T_N)$.*
The following equality holds true:

$$T_i = \min(g_i,\gamma_i) = \max(l_{i+1},\lambda_{i+1}),\ i=1,\dots,N-1 \tag{3.4}$$

and on Assumption 2(a) $\delta_N := \sup_{i\in\{1,\dots,N\}}(T_i-T_{i-1}) = \mathcal{O}\left(N^{-2/3-\alpha}\right)$ holds. Analogously, $\delta_N = \mathcal{O}\left(N^{-8/9-\alpha}\right)$ on Assumption 2(b) holds.

Proof. Assume without loss of generality $g_i \le \gamma_i$ for an arbitrarily fixed $i\in\{1,\dots,N-1\}$. Taking Algorithm 3.1 into account, we proof that (3.4) holds true.
If $g_i<\gamma_i$, then the observation times γ_i and $g_{i,+} := \min\left(t_k\in\mathcal{T}^X | t_k > g_i\right)$ are compared in the $(i+1)$th step of the synchronization Algorithm 3.1 and $g_{i,+} = \min\left(t_k | t_k \in \mathcal{H}^{i+1}\right)$ holds true. Thus, $g_i = l_{i+1}$ and (3.4) holds true. We remark that in this case $\gamma_i\in\mathcal{G}^{i+1}$ and thus $\gamma_i > \lambda_{i+1} = \gamma_{i,-} := \max\left(\tau_k\in\mathcal{T}^Y|\tau_k<\gamma_i\right) \ge \gamma_{i-1}$.
If $g_i=\gamma_i$, then the observation times $g_{i,+}$ and $\gamma_{i,+}$ are compared in the $(i+1)$th step of Algorithm 3.1 and $l_{i+1}=\lambda_{i+1}=g_i=\gamma_i$ what implies (3.4).
Equation (3.4) does not hold true for $i=0,N$ and $T_0 = t_0\wedge\tau_0$ because of our definition that $l_0=t_0$ and $\lambda_0=\tau_0$.
Although consecutive maxima g_i of the sets $\mathcal{H}^i$ and γ_i of the sets $\mathcal{G}^i$, respectively, can be equal, $T_i > T_{i-1}$ holds for all $i\in\{1,\dots,N\}$ because $g_{i+1}=g_i$ implies that $\gamma_{i+1}>\gamma_i$ and $\gamma_{i+1}=\gamma_i$ implies that $g_{i+1}>g_i$. Hence, the set $\mathcal{T}^{syn}$ induces a partition of the time span $[0,T]$. □

The times $T_i, i=0,\dots,N$ defined through (3.4) are the refresh times that have been mentioned in the last section. We use Proposition 3.2.1 to split the error of the estimator

(3.2) for the integrated covariance $\langle X,Y\rangle_T$ in two asymptotically uncorrelated parts. The error of the estimator (3.2) can be written

$$\sum_{i=1}^{N}\left(X_{g_i}-X_{l_i}\right)\left(Y_{\gamma_i}-Y_{\lambda_i}\right)-\int_0^T \rho_t\sigma_t^X\sigma_t^Y\,dt = D_T^N + A_T^N$$

where

$$D_T^N := \sum_{i=1}^{N}\left(\left(X_{T_i}-X_{T_{i-1}}\right)\left(Y_{T_i}-Y_{T_{i-1}}\right)-\int_{T_{i-1}}^{T_i}\rho_t\sigma_t^X\sigma_t^Y\,dt\right) \tag{3.5}$$
$$-\int_0^{t_0\wedge\tau_0}\rho_t\sigma_t^X\sigma_t^Y\,dt-\int_{t_n\wedge\tau_m}^{T}\rho_t\sigma_t^X\sigma_t^Y\,dt$$

is the discretization error of a realized covariance estimator evaluated with synchronous observations at the times $T_i, i=0,\ldots,N$, which is the closest synchronous approximation to the asynchronous sampling scheme, and

$$\begin{aligned}A_T^N =& \sum_{i=1}^{N}\left(Y_{\gamma_i}-Y_{\lambda_i}\right)\left(X_{g_i}-X_{T_i}\right)\mathbb{1}_{\{T_i=\gamma_i\}}+\left(Y_{T_i}-Y_{T_{i-1}}\right)\left(X_{T_{i-1}}-X_{l_i}\right)\mathbb{1}_{\{T_{i-1}=\lambda_i\}}\\ &+\sum_{i=1}^{N}\left(X_{T_i}-X_{l_i}\right)\left(Y_{\gamma_i}-Y_{T_i}\right)\mathbb{1}_{\{T_i=g_i\}}+\left(X_{T_i}-X_{T_{i-1}}\right)\left(Y_{T_{i-1}}-Y_{\lambda_i}\right)\mathbb{1}_{\{T_{i-1}=l_i\}}\end{aligned} \tag{3.6}$$

is the remaining additional error due to the lack of synchronicity.

Proposition 3.2.2. *The Brownian parts of A_T^N and D_T^N are uncorrelated. This means, that if we assume the drift terms to be identically zero in Assumption 1, A_T^N and D_T^N are uncorrelated. If the drift terms are non-zero, A_T^N and D_T^N are asymptotically uncorrelated.*

Proof. A_T^N and D_T^N are both centred. If Assumption 1 holds with $\mu_t^X=\mu_t^Y\equiv 0$, the expectation of the product of A_T^N and D_T^N is zero, since the previous- and next-tick interpolated increments in (3.6) are centred and uncorrelated to the other three factors in each addend of the inner sums.
If we allow for non-zero drift terms, Assumption 1 and Assumption 2 ensure that the increments over time intervals due to the drift induce terms at most of order δ_N in probability by products of drift terms and at most of order $\delta_N^{1/2}$ in probability by products of drift and Brownian terms in the overall correlation. □

In Figure 3.4 the observation times $\tau_j, j=0,\ldots,11$ of Y for our Example 3.2 from the last section are plotted against the observation times $t_i, i=0,\ldots,11$ of X. The dashed lines intersect for synchronous observation times $t_0=\tau_0, t_3=\tau_3$ and $t_{10}=\tau_{10}$ on the diagonal of the square in Figure 3.4. A similar visualization of the realized covariance estimator for synchronous and equidistant data would yield coextensive squares around the diagonal, over which multiplied increments are summed up. Refresh times are (in general) not equidistant but provide a synchronous realized covariance estimator as an approximation. The Hayashi-Yoshida estimator (3.2) is the sum of products of increments with overlapping observation time instants. The relation to the synchronous approximation D_T^N is that we have next-tick interpolations and previous-tick interpolations to the times $T_i, i = 0, \ldots, 8$ and take increments from previous-tick interpolated values to next-tick interpolated values. The time instants of D_T^N are visualized for our example in Figure 3.4. The previous- and next-tick interpolations are illustrated in Figure 3.5. The products of time instants leading to the error A_T^N are illustrated in the same picture by the grey rectangles. As can be seen for the example in Figure 3.5, A_T^N is the sum of the errors by the ith next-tick interpolation multiplied with the increments of the other process over $[\min(l_i, \lambda_i), T_i]$ and the sum of the errors of the ith previous-tick interpolation multiplied with the increments of the other process over $[T_{i-1}, T_i]$. The sum of the increments over the squares in Figure 3.4, D_T^8 for our example, and the grey rectangles in Figure 3.5, A_T^8 for our example, is the Hayashi-Yoshida estimator evaluated at the end of the last section.

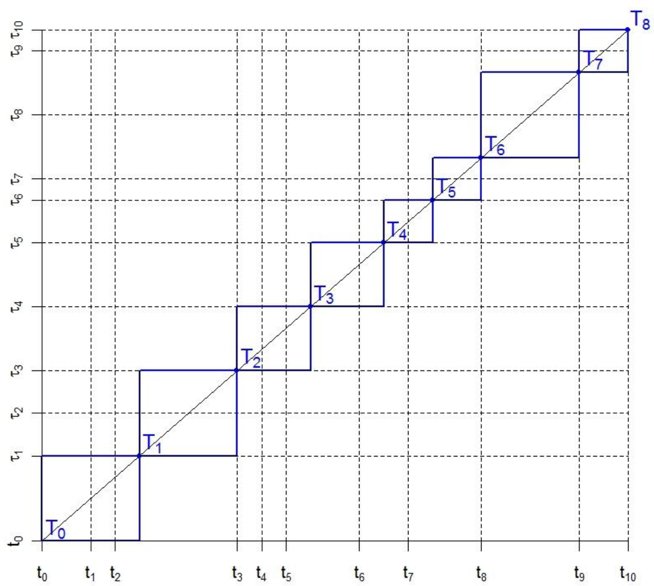


Figure 3.4: Illustration of the synchronous approximation for our example.

Definition 3.2.3 (quadratic (co-)variations of time). *For any $N \in \mathbb{N}$ let $T_i^{(N)}$, $i = 0, \ldots, N$ be the times from the partition of $[0, T]$ defined in* (3.4) *above and $g_i^{(N)}, \gamma_i^{(N)}, l_i^{(N)}$, $\lambda_i^{(N)}$ the corresponding observation times defined above in the estimator* (3.2)*. T/N is the mean of the time instants $\Delta T_i^{(N)} = T_i^{(N)} - T_{i-1}^{(N)}$, $i = 1, \ldots, N$. Define the following*

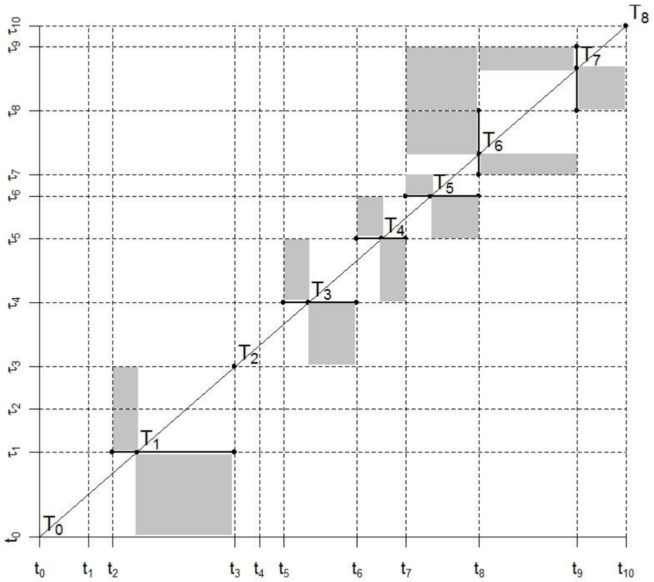


Figure 3.5: Illustration of the next- and previous-tick interpolated values and the error due to non-synchronicity for our example.

sequences of functions

$$G^N(t) = \frac{N}{T} \sum_{T_i^{(N)} \le t} \left(\Delta T_i^{(N)}\right)^2 \ , \tag{3.7a}$$

$$\begin{aligned} F^N(t) = \frac{N}{T} \sum_{T_{i+1}^{(N)} \le t} & (T_i^{(N)} - \lambda_i^{(N)})(g_i^{(N)} - T_i^{(N)}) + \left(T_i^{(N)} - l_i^{(N)}\right)\left(\gamma_i^{(N)} - T_i^{(N)}\right) \\ & + \Delta T_{i+1}^{(N)}\left(T_i^{(N)} - l_{i+1}^{(N)}\right) + \Delta T_{i+1}^{(N)}\left(T_i^{(N)} - \lambda_{i+1}^{(N)}\right) \ , \end{aligned} \tag{3.7b}$$

$$H^N(t) = \frac{N}{T} \sum_{T_{i+1}^{(N)} \le t} \left(T_i^{(N)} - l_{i+1}^{(N)}\right)\left(g_i^{(N)} - T_i^{(N)}\right) + \left(T_i^{(N)} - \lambda_{i+1}^{(N)}\right)\left(\gamma_i^{(N)} - T_i^{(N)}\right) \ , \tag{3.7c}$$

for $t \in [0, T]$ that we call sequences of quadratic (co-)variations of times.

A stable central limit theorem for the estimation error is deduced on the assumption that the sequences defined by (3.7a), (3.7b) and (3.7c) converge pointwise and the sequences of difference quotients uniformly:

Assumption 3.1 (asymptotic quadratic (co-)variation of times). *Assume that for the sequences of sampling schemes and the times $T_i^{(N)}, g_i^{(N)}, \gamma_i^{(N)}, l_i^{(N)}, \lambda_i^{(N)}$ and the sequences of quadratic (co-) variations of times $G^N(t), F^N(t), H^N(t)$ defined in Definition 3.2.3 the following holds true:*

(i) $G^N(t) \to G(t)$, $F^N(t) \to F(t)$, $H^N(t) \to H(t)$ *as* $N \to \infty$, *where* $G(t), F(t), H(t)$ *are continuously differentiable functions on* $[0, T]$.

(ii) For any null sequence (h_N), $h_N = \mathcal{O}\left(N^{-1}\right)$

$$\frac{G^N(t+h_N) - G^N(t)}{h_N} \to G'(t) \tag{3.8a}$$

$$\frac{F^N(t+h_N) - F^N(t)}{h_N} \to F'(t) \tag{3.8b}$$

$$\frac{H^N(t+h_N) - H^N(t)}{h_N} \to H'(t) \tag{3.8c}$$

uniformly on [0,T] as $N \to \infty$.

Assumption 3.1 is necessary to ensure that the sequence of variances of the estimator (3.2) converges as $n, m \to \infty$. The derivative of the asymptotic quadratic variation of refresh times (3.8a) will appear in the asymptotic variance of the discretization error D_T^N, since refresh times are (in general) not equidistant. For $\Delta T_i^{(N)} = T/N$ for all $i \in \{0, \ldots, N\}$, $G'(t) = \mathbb{1}_{[0,T]}$ holds true.
The uniform convergence of the difference quotients defined by (3.8b) and (3.8c) are necessary to ensure that the sequence of variances of A_T^N converges as $N \to \infty$. The assumptions imposed by (3.8a)-(3.8c) are weaker than assuming convergence of the joint sampling design of $\left(\mathcal{T}^{X,n}, \mathcal{T}^{Y,m}\right)$ and are not very restrictive. They hold true whenever the sequences of sampling schemes tend to a certain state of asynchronicity or have a uniform behaviour of non-synchronicity in the limit as $n, m \to \infty$. For homogeneous sampling schemes these (co-)variations of time converge to linear limiting functions.

Example:
Consider the synchronous equidistant sampling schemes with $N = n = m$ and $t_i^{(n)} = \tau_j^{(n)} = i/n, i = 0, \ldots, n$. The left-hand side of Figure 3.6 shows the quadratic (co)variations of time G^N, F^N and H^N for $N = 30000$. F^N and H^N are identically zero since there are no asynchronous observations. The function G^N is a step function that will tend to the identity on $[0, T]$ as $N \to \infty$.
Next, we consider a situation of completely non-synchronous sampling schemes that will be called intermeshed sampling. This originates from the complete synchronous equidistant one by shifting one time-scale half a time instant $1/2N$. In this case the synchronous approximation is still equidistant with instants $1/N$ and, hence, G is the

identity function. F and H are linear limiting functions with slope 1 and 1/4, respectively. The functions G^N, F^N, H^N are illustrated in Figure 3.6 on the right-hand side.
The other functions that can be seen in Figure 3.6 are the so-called degrees of regularity of non-synchronicity for the two examples. This notion that will affect the asymptotics of the combined estimator in Chapter 4 is introduced below in Definition 4.2.1. In Chapter 5

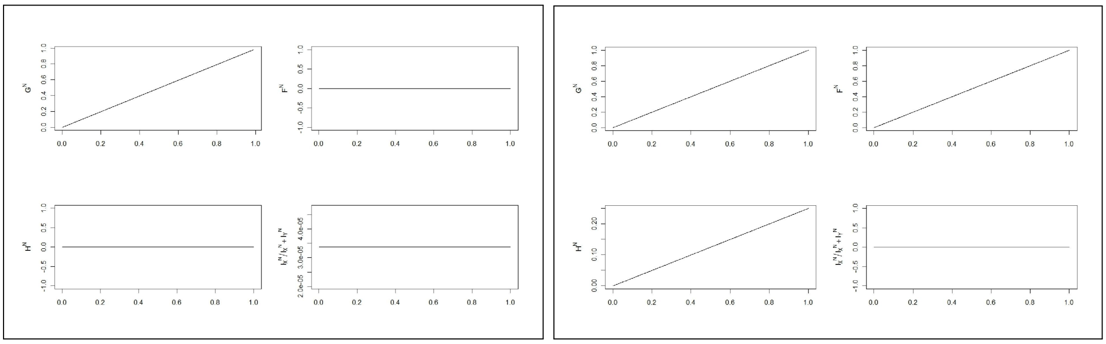


Figure 3.6: Quadratic (Co-)variations of time for synchronous equidistant (left) and intermeshed (right) sampling.

we show that for an important special case, independent homogeneous Poisson sampling, (3.8a)-(3.8c) are fulfilled when replacing deterministic convergence by convergence in probability. Furthermore, the stochastic limits $G'(t), F'(t), H'(t)$ are calculated explicitly and are again constant on $[0, T]$. For data applications one can calculate easily empirical versions $\tilde{G}'_{n,m}(t), \tilde{F}'_{n,m}(t), \tilde{H}'_{n,m}(t)$ of $G'(t), F'(t), H'(t)$ and use those as estimators for (3.8a)-(3.8c).
The main result of this chapter is Theorem 3.1. This result serves as preparation to prove the stable limit theorem of the generalized multiscale estimator in Theorem 4.1 in Chapter 4 but, furthermore, it gives insight into the asymptotic distribution of the Hayashi-Yoshida estimator. It improves on the asymptotic normality result in Hayashi and Yoshida [2008], since the weak convergence is stable in the setting where we allow for random correlation, drift and volatility processes. Apart from that, the representation of the asymptotic variance using (3.8a)-(3.8c) differs from that in Hayashi and Yoshida [2008] by the decomposition in (3.5) and (3.6) and the notion of (co-)variations of times.

Theorem 3.1. *The estimation error of* (3.2) *converges on the Assumptions 1, 2(a) and 3.1 stably in law to a centred, mixed Gaussian distribution:*

$$\sqrt{N}\left(\sum_{i=1}^{N}\left(X_{g_i}-X_{l_i}\right)\left(Y_{\gamma_i}-Y_{\lambda_i}\right)-\langle X\,,\,Y\rangle_T\right)\overset{st}{\rightsquigarrow}\mathbf{N}\left(0\,,\,v_{D_T}+v_{A_T}\right)\ , \tag{3.9}$$

with the asymptotic variance

$$\underbrace{T\int_0^T G'(t)\left(\sigma_t^X\sigma_t^Y\right)^2\left(\rho_t^2+1\right)dt}_{=v_{D_T}}+\underbrace{T\int_0^T\left(F'(t)\left(\sigma_t^X\sigma_t^Y\right)^2+2H'(t)\left(\rho_t\sigma_t^X\sigma_t^Y\right)^2\right)dt}_{=v_{A_T}}$$

where the two addends come from the asymptotic variances of D_T^N and A_T^N, respectively.

3.3 Proof of the stable central limit theorem

3.3.1 Discretization error of the synchronous approximation

Proposition 3.3.1. *On the Assumptions 1, 2(a) and* (3.8a) *the discretization error of the closest synchronous approximation converges stably in law to a centred mixed Gaussian distribution:*

$$\sqrt{\frac{N}{T}}D_T^N \overset{st}{\rightsquigarrow} \mathbf{N}\left(0\,,\int_0^T G'(t)\left(\sigma_t^X\sigma_t^Y\right)^2(\rho_t^2+1)\,dt\right)\,. \tag{3.10}$$

Proof. First note that on Assumption 1, by Girsanov's Theorem 1.2 we may without loss of generality further suppose that $\mu_t^X=\mu_t^Y=0$ identically. This is possible with the stability of the weak convergence which is ensured by asymptotic independence between the limiting normal distribution and the martingale parts of the observed processes using Jacod's Theorem 1.6 (cf. the discussion in Subsection 1.1.2 and Section 1.2).
As before, we often omit the superscripts (N) for the sampling times to increase the readability.
Let M_t and L_t be the continuous martingales $L_t=\int_0^t\sigma_s^X\,dW_s^X$, $M_t=\int_0^t\sigma_s^Y\,dW_s^Y$ with standard Brownian motions W^X,W^Y with quadratic covariation $\langle W^X,W^Y\rangle_t=\int_0^t\rho_s\sigma_s^X\sigma_s^Y\,ds$, that represent the transformed martingale processes, and $L_i=\int_0^{T_i}\sigma_s^X dW_s^X$, $M_i=\int_0^{T_i}\sigma_s^Y\,dW_s^Y$.

Proposition 3.3.2. *On the same Assumptions as in Proposition 3.3.1, the process $\mathcal{D}_t^N$ defined by*

$$\mathcal{D}_t^N:=\sqrt{\frac{N}{T}}\sum_{T_i^{(N)}\le t}(L_i-L_{i-1})(M_i-M_{i-1})-\int_0^t\rho_s\sigma_s^X\sigma_s^Y\,ds$$

for $0\le t\le T$ converges as $N\to\infty$ stably in law:

$$\mathcal{D}_t^N\overset{st}{\rightsquigarrow}\int_0^t\sqrt{v_{\mathcal{D}_s}}dW_s^{\perp} \tag{3.11}$$

where $W^{\perp}$ is a Brownian motion independent of $\mathcal{F}$ and

$$v_{\mathcal{D}_s}=G'(s)(\sigma_s^X\sigma_s^Y)^2(\rho_s^2+1)\,. \tag{3.12}$$

Proof. We will prove the proposition by application of Jacod's Theorem 1.6. It would also be possible to use the discrete-time version of this Theorem 1.3.5 which will be applied in the next subsection. Here, we prefer to consider the continuous-time version to gain a better understanding of the key elements that lead to the stable central limit

theorem in Proposition 3.3.2.
Using the definition of the quadratic covariation process of martingales (or integration by parts formula) we can find an illustration of the discretization error by a sum of stochastic integrals and an asymptotically negligible term:

$$\begin{aligned}
&\sum_{T_i^{(N)}\le t} \left(L_{T_i}-L_{T_{i-1}}\right)\left(M_{T_i}-M_{T_{i-1}}\right) = \sum_{T_i^{(N)}\le t} \left(L_i-L_{i-1}\right)\left(M_i-M_{i-1}\right)\\
&= \sum_{T_i^{(N)}\le t} \left(L_iM_i - L_iM_{i-1} - M_iL_{i-1} + L_{i-1}M_{i-1}\right)\\
&= \sum_{T_i^{(N)}\le t} \Big(\int_{T_{i-1}}^{T_i} L_s dM_s + \int_{T_{i-1}}^{T_i} M_s dL_s + \Delta\langle L,M\rangle_{T_i}\\
&\qquad - M_{i-1}(L_i-L_{i-1}) - L_{i-1}(M_i-M_{i-1})\Big)\\
&= \langle L,M\rangle_{\mathfrak{T}(t)} - \langle L,M\rangle_{T_0} + \sum_{T_i^{(N)}\le t}\left(\int_{T_{i-1}}^{T_i}(L_s-L_{i-1})dM_s + \int_{T_{i-1}}^{T_i}(M_s-M_{i-1})dL_s\right)
\end{aligned}$$

where we denote $\mathfrak{T}(t) := \max_i \left(T_i^{(N)} \le t\right)$.
Thus, we obtain

$$\sqrt{\frac{N}{T}}\mathcal{D}_t^N = \sqrt{\frac{N}{T}}\sum_{T_i^{(N)}\le t}\left(\int_{T_{i-1}}^{T_i}(L_s-L_{i-1})dM_s + \int_{T_{i-1}}^{T_i}(M_s-M_{i-1})dL_s\right) + \mathcal{O}_p\left(\sqrt{\frac{N}{T}}\right) ,$$

since $\langle L,M\rangle_{\mathfrak{T}(t)} - \langle L,M\rangle_{T_0} = \langle L,M\rangle_t + \mathcal{O}_p(1)$. Consider the centred continuous martingale

$$\begin{aligned}
\phi_\tau^{(N)} := \sqrt{\frac{N}{T}}\Big(&\sum_{T_i^{(N)}\le t}\left(\int_{T_{i-1}}^{T_i}(L_s-L_{i-1})dM_s + \int_{T_{i-1}}^{T_i}(M_s-M_{i-1})dL_s\right)\\
&+ \int_{\mathfrak{T}(t)}^{\tau}(L_s-L_{\mathfrak{T}(t)})dM_s + \int_{\mathfrak{T}(t)}^{\tau}(M_s-M_{\mathfrak{T}(t)})dL_s\Big) \ , \ \tau\in[\mathfrak{T}(t),t] \ .
\end{aligned}$$

We calculate the corresponding quadratic variation process at time t:

$$\begin{aligned}
\langle\phi^{(N)}\rangle_t &= \frac{N}{T}\left\langle \sum_{T_i^{(N)}\le t}\left(\int_{T_{i-1}}^{T_i}(L_s-L_{i-1})dM_s + \int_{T_{i-1}}^{T_i}(M_s-M_{i-1})dL_s\right)\right\rangle_t\\
&\quad + \langle\phi^{(N)}\rangle_t - \langle\phi^{(N)}\rangle_{\mathfrak{T}(t)}\\
&= \frac{N}{T}\sum_{T_i^{(N)}\le t}\Big(\int_{T_{i-1}}^{T_i}(L_s-L_{i-1})^2 d\langle M\rangle_s + \int_{T_{i-1}}^{T_i}(M_s-M_{i-1})^2 d\langle L\rangle_s\\
&\quad + 2\int_{T_{i-1}}^{T_i}(L_s-L_{i-1})(M_s-M_{i-1})d\langle M,L\rangle_s\Big) + \langle\phi^{(N)}\rangle_t - \langle\phi^{(N)}\rangle_{\mathfrak{T}(t)}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{N}{T} \sum_{T_i^{(N)} \le t} \Big(\int_{T_{i-1}}^{T_i} \langle L - L_{i-1} \rangle_s d\langle M \rangle_s + \int_{T_{i-1}}^{T_i} \langle M - M_{i-1} \rangle_s d\langle L \rangle_s \\
&\quad + 2 \int_{T_{i-1}}^{T_i} \langle L - L_{i-1} \rangle_s \langle M - M_{i-1} \rangle_s d\langle M, L \rangle_s \Big) + \mathcal{o}_p(1) \\
&= \frac{N}{T} \sum_{i=1}^{N} \Big(\int_{T_{i-1}}^{T_i} d\left(\langle L - L_{i-1} \rangle_s \langle M - M_{i-1} \rangle_s \right) \\
&\quad + 2 \int_{T_{i-1}}^{T_i} \langle L - L_{i-1} \rangle_s \langle M - M_{i-1} \rangle_s d\langle M, L \rangle_s \Big) + \mathcal{o}_p(1) \\
&= \frac{N}{T} \sum_{T_i^{(N)} \le t} \left(\int_{T_{i-1}}^{T_i} \left(\sigma_s^X \right)^2 ds \int_{T_{i-1}}^{T_i} \left(\sigma_s^Y \right)^2 ds \right) + \left(\int_{T_{i-1}}^{T_i} \rho_s \sigma_s^X \sigma_s^Y \, ds \right)^2 + \mathcal{o}_p(1) \\
&= \frac{N}{T} \sum_{T_i^{(N)} \le t} \left(\left((\overline{\rho \sigma^X \sigma^Y})_i \right)^2 (\Delta T_i)^2 + \left((\overline{\sigma^X})_i (\overline{\sigma^Y})_i \right)^2 (\Delta T_i)^2 \right) + \mathcal{o}_p(1) \\
&= \sum_{T_i^{(N)} \le t} \left(\frac{\left(G^{(N)}(T_i) - G^{(N)}(T_{i-1}) \right)}{\Delta T_i} \left(\sigma_{T_{i-1}}^X \sigma_{T_{i-1}}^Y \right)^2 \left(1 + {\rho_{T_{i-1}}}^2 \right) \Delta T_i \right) + \mathcal{o}_p(1) \\
&\xrightarrow{p} \int_0^t G'(s)(\rho_s^2 + 1)(\sigma_s^X \sigma_s^Y)^2 \, ds
\end{aligned}
$$

The third equality including a remainder term of $\mathcal{o}_p(1)$ is proved in the next lemma. In this calculation we further used the integration by parts formula (1.1.3) in the following step and then the change of variables Theorem 1.1.5 for the integrals with quadratic covariation integrators that are of finite variation. The second last equality is an application of the mean value theorem (the volatility and the correlation processes are continuous and thus also bounded on compact sets) where the constants $(\overline{\sigma^X})_i$, $(\overline{\sigma^Y})_i$ and $(\overline{\rho \sigma^X \sigma^Y})_i$ come from. The Riemann sum converges and using Definition 3.2.3 and Assumption 3.1, the convergence in probability of the quadratic variation to $\int_0^t G'(s)(\rho_s^2 + 1)(\sigma_s^X \sigma_s^Y)^2 \, ds = \int_0^t v_{\mathcal{D}_s}$ follows.

Lemma 3.3.3. *Using the notation as above the following relations hold true:*

$$
\sum_{T_i^{(N)} \le t} \int_{T_{i-1}}^{T_i} \left((L_s - L_{i-1})^2 - \langle L - L_{i-1} \rangle_s \right) d\langle M - M_{i-1} \rangle_s = \mathcal{o}_p(1) \tag{3.13a}
$$

$$
\sum_{T_i^{(N)} \le t} \int_{T_{i-1}}^{T_i} \left((M_s - M_{i-1})^2 - \langle M - M_{i-1} \rangle_s \right) d\langle L - L_{i-1} \rangle_s = \mathcal{o}_p(1) \tag{3.13b}
$$

$$\sum_{T_i^{(N)}\leq t}\int_{T_{i-1}}^{T_i}((M_s-M_{i-1})(L_s-L_{i-1})-\langle M-M_{i-1},L-L_{i-1}\rangle_s)\,d\langle M,L\rangle_s=\mathcal{o}_p(1) \tag{3.13c}$$

$$\frac{N}{T}\sum_{T_i^{(n)}\leq t}(\Delta T_i)^2\left((\overline{\rho\sigma^X\sigma^Y})_i^2+(\overline{\sigma^X})_i^2(\overline{\sigma^Y})_i^2-\left(\rho_{T_{i-1}}\sigma^X_{T_{i-1}}\sigma^Y_{T_{i-1}}\right)^2-\left(\sigma^X_{T_{i-1}}\sigma^Y_{T_{i-1}}\right)^2\right)=\mathcal{o}_p(1) \tag{3.13d}$$

Proof. The proofs of (3.13a) and (3.13b) are completely analogous. We prove (3.13a). By Itô's formula

$$(L_s-L_{i-1})^2=2\int_{T_{i-1}}^{s}(L_r-L_{i-1})dL_r+\langle L-L_{i-1}\rangle_s$$

holds. The left-hand side of (3.13a) equals

$$\sum_{T_i^{(N)}\leq t}\left(\int_{T_{i-1}}^{T_i}\left(2\int_{T_{i-1}}^{s}(L_r-L_{i-1})dL_r\right)d\langle M-M_{i-1}\rangle_r\right)$$
$$=\sum_{T_i^{(N)}\leq t}\left(2\int_{T_{i-1}}^{T_i}(L_s-L_{i-1})(\langle M-M_{i-1}\rangle_{T_i})dL_s-2\int_{T_{i-1}}^{T_i}(L_s-L_{i-1})(\langle M-M_{i-1}\rangle_s)dL_s\right)$$

by application of the integration by parts formula (1.1.3) in the way

$$Z_{T_i}\langle M-M_{i-1}\rangle_{T_i}=\int_0^{T_i}Z_t d\langle M-M_{i-1}\rangle_t+\int_0^{T_i}\langle M-M_{i-1}\rangle_t dZ_t$$

with $Z_t:=\int_{T_{i-1}}^t 2(L_s-L_{i-1})dL_s$ for $T_{i-1}\leq t\leq T$ to the addends. Therefore, we can write the left-hand side of (3.13a) in the way $\mathcal{M}_1^{(N)}+\mathcal{M}_2^{(N)}$ with two centred continuous martingales $\mathcal{M}_1^{(N)},\mathcal{M}_2^{(N)}$ defined in the fashion of $\phi^{(N)}$ above and calculate the quadratic covariation processes at time t:

$$\langle\mathcal{M}_2^{(N)}\rangle_t=4\sum_{T_i^{(N)}\leq t}\left(\int_{T_{i-1}}^{T_i}(L_s-L_{i-1})^2(\langle M-M_{i-1}\rangle_s)^2 d\langle L\rangle_s\right)+\mathcal{o}_p(1)$$
$$\leq 4\max_i\sup_{s\in(T_{i-1},T_i]}(L_s-L_{i-1})^2\max_i\sup_{s\in(T_{i-1},T_i]}\langle M-M_{i-1}\rangle_s^2\sum_{T_i^{(N)}\leq t}\int_{T_{i-1}}^{T_i}d\langle L\rangle_s+\mathcal{o}_p(1)\,.$$

The first addend is up to a logarithmic factor $\mathcal{O}_p(\delta_N^3)$ and hence $\mathcal{M}_2^{(N)}=\mathcal{o}_p(1)$ on Assumption 2. That $\mathcal{M}_1^{(N)}=\mathcal{o}_p(1)$ is proved analogously. This proves (3.13a).
The strategy to prove (3.13c) follows the same approach. For the sake of completeness

we give the first part of the proof in the following. We begin with the equation

$$(L_s - L_{i-1})(M_s - M_{i-1}) = \int_{T_{i-1}}^{s} (L_r - L_{i-1})d(M - M_{i-1})_r \\ + \int_{T_{i-1}}^{s} (M_r - M_{i-1})d(L - L_{i-1})_r + \langle L - L_{i-1}, M - M_{i-1}\rangle_s$$

and apply the integration by parts formula as above with $Z_t = \int_{T_{i-1}}^{t}(L_s - L_{i-1})d(M - M_{i-1})_s + \int_{T_{i-1}}^{t}(M_s - M_{i-1})d(L - L_{i-1})_s$ for $T_{i-1} \le t \le T_i$. This yields for the left-hand side of (3.13c)

$$\sum_{T_i^{(N)} \le t} \left(\int_{T_{i-1}}^{T_i} \left(\int_{T_{i-1}}^{s} (L_r - L_{i-1})d(M - M_{i-1})_r \right) d\langle M, L\rangle_s \right. \\ \left. + \int_{T_{i-1}}^{T_i} \left(\int_{T_{i-1}}^{s} (M_r - M_{i-1})d(L - L_{i-1})_r \right) d\langle M, L\rangle_s \right)$$
$$= \sum_{T_i^{(N)} \le t} \left(\int_{T_{i-1}}^{T_i} (L_s - L_{i-1})(\langle M, L\rangle_{T_i} - \langle M, L\rangle_{T_{i-1}})d(M - M_{i-1})_s \right. \\ + \int_{T_{i-1}}^{T_i} (M_s - M_{i-1})(\langle M, L\rangle_{T_i} - \langle M, L\rangle_{T_{i-1}})d(L - L_{i-1})_s \\ \left. - \int_{T_{i-1}}^{T_i} (L_s - L_{i-1})\langle M, L\rangle_s d(M - M_{i-1})_s - \int_{T_{i-1}}^{T_i} (M_s - M_{i-1})\langle M, L\rangle_s d(L - L_{i-1})_s \right)$$

Now one can proceed as for (3.13a) to prove (3.13c).
We complete the proof of the convergence of the quadratic variation with the proof of approximation (3.13d). Denote $\overline{(\rho\sigma^X\sigma^Y)}_i^2 = \widetilde{(\sigma^X)^2}_i \cdot \widetilde{(\sigma^Y)^2}_i \cdot \widetilde{(\rho)^2}_i$ to distinguish between the values from the application of the mean value theorems to the two different addends. An upper bound of the left-hand side of (3.13d) can be found by elementary algebra and the triangle inequality for the absolute value:

$$\frac{N}{T} \sum_{T_i^{(N)} \le t} (\Delta T_i)^2 \left((\widetilde{\rho\sigma^X\sigma^Y})_i^2 + (\overline{\sigma^X})_i^2(\overline{\sigma^Y})_i^2 - \left(\left(\rho_{T_{i-1}}\sigma^X_{T_{i-1}}\sigma^Y_{T_{i-1}}\right)^2 + \left(\sigma^X_{T_{i-1}}\sigma^Y_{T_{i-1}}\right)^2 \right) \right)$$
$$\le \frac{N}{T} \sum_{T_i^{(N)} \le t} (\Delta T_i)^2 \left(\widetilde{(\sigma^X)^2}_i \widetilde{(\sigma^Y)^2}_i \left| \widetilde{(\rho)^2}_i - \rho^2_{T_{i-1}} \right| + \widetilde{(\sigma^Y)^2}_i \rho^2_{T_{i-1}} \left| \widetilde{(\sigma^X)^2}_i - (\sigma^X_{T_{i-1}})^2 \right| \right. \\ + \rho^2_{T_{i-1}}(\sigma^X_{T_{i-1}})^2 \left| \widetilde{(\sigma^Y)^2}_i - (\sigma^Y_{T_{i-1}})^2 \right| + (\overline{\sigma^Y})_i^2 \left| (\overline{\sigma^X})_i^2 - (\sigma^X_{T_{i-1}})^2 \right| \\ \left. + (\sigma^X_{T_{i-1}})^2 \left| (\overline{\sigma^Y})_i^2 - (\sigma^Y_{T_{i-1}})^2 \right| \right) = \mathcal{O}_p(1) \ .$$

□

For the martingales $\phi^{(N)}$ there are representations as time-changed Brownian motions

$B^{(DDS,N)}_{\langle\phi^{(N)}\rangle_t} = \phi^{(N)}_t$ by the Dambis-Dubins-Schwarz Theorem 1.3. The sequence of martingales $\phi^{(N)}$ or associated time-changed Dambis-Dubins-Schwarz Brownian motions converges weakly to a limiting Brownian motion by Theorem 1.4. This convergence is stable and the limiting Brownian motion is defined on an orthogonal extension of the original probability space. To obtain the stable convergence result, we are left to verify conditions (1.3a) and (1.3b) of Jacod's Theorem 1.6.
Consider the quadratic covariation process of $\phi^{(N)}$ and the reference martingale L

$$\langle L,\phi^{(N)}\rangle_t = \sqrt{\frac{N}{T}}\sum_{T_i^{(N)}\le t}\left(\int_{T_{i-1}}^{T_i}(L_s-L_{i-1})d\langle M,L\rangle_s+\int_{T_{i-1}}^{T_i}(M_s-M_{i-1})d\langle L\rangle_s\right)+\mathcal{O}_p(1)\,.$$

The term of smaller order than 1 in probability comes from the increment of the covariation process on $[\mathfrak{T}(t),t]$. As before, this equality holds true for all t, since for $t<T_1$ the covariation is $\mathcal{O}_p(1)$. Integration by parts yields:

$$\begin{aligned}\langle L,\phi^{(N)}\rangle_t = \sqrt{\frac{N}{T}}\sum_{T_i^{(N)}\le t}\Bigg[&(\langle M,L\rangle_{T_i}-\langle M,L\rangle_{T_{i-1}})(L_i-L_{i-1})-\int_{T_{i-1}}^{T_i}\langle M,L\rangle_s d(L_s-L_{i-1})\\ &+(\langle L\rangle_{T_i}-\langle L\rangle_{T_{i-1}})(M_i-M_{i-1})-\int_{T_{i-1}}^{T_i}\langle L\rangle_s d(M_s-M_{i-1})\Bigg]\,.\end{aligned}$$

It remains to show that this term converges to zero in probability. The term is centred and using Itô isometry we find the following upper bound for the second moment:

$$\begin{aligned}\mathbb{E}\left[\left(\langle L,\phi^{(N)}\rangle_t\right)^2\right] \le 2\frac{N}{T}\mathbb{E}\Bigg[&\sum_{T_i^{(N)}\le t}\Big((\langle M,L\rangle_{T_i}-\langle M,L\rangle_{T_{i-1}})^2(L_i-L_{i-1})^2\\ &+(\langle L\rangle_{T_i}-\langle L\rangle_{T_{i-1}})^2(M_i-M_{i-1})^2\Big)\\ &+\max_{i\in\{1,\dots,N\}}\sup_{s\in(T_{i-1},T_i]}(\langle M,L\rangle_s-\langle M,L\rangle_{T_{i-1}})^2\sum_i\int_{T_{i-1}}^{T_i}d\langle L-L_{i-1}\rangle_t\\ &+\max_{i\in\{1,\dots,N\}}\sup_{s\in(T_{i-1},T_i]}(\langle L\rangle_s-\langle L\rangle_{T_{i-1}})^2\sum_i\int_{T_{i-1}}^{T_i}d\langle M-M_{i-1}\rangle_t\Bigg]\\ =\mathcal{O}\left(N\delta_N^2\right)\,.&\end{aligned}$$

The term is bounded by a constant times $N\delta_N^2$ since squared increments, cross products of increments and increments of the quadratic (co-)variations of L and M over time instants $\Delta T_i^{(N)}$ are bounded by $\Delta T_i^{(N)}$ times a constant. To sums with products of time instants we can apply Hölder's inequality with the supremum norm to obtain upper bounds. There are at most order δ_N^{-1} time instants $\Delta T_i^{(N)}$ of order $\sup_i \Delta T_i^{(N)}=\delta_N$ since $\sum_i \Delta T_i^{(N)}\le T$ and the time span T is fixed.

Hence, $\langle L, \phi^{(N)} \rangle_t = \mathcal{O}_p(1) \quad \forall t \in [0,T]$. With the same strategy $\langle M, \phi^{(N)} \rangle_t = \mathcal{O}_p(1) \quad \forall t \in [0,T]$ can be proven.
For every bounded $\mathcal{F}_t$-martingale $L^{\perp}$ satisfying $\langle L, L^{\perp} \rangle \equiv 0$ the covariation

$$\langle L^{\perp}, \phi^{(N)} \rangle_t = \frac{N}{T} \sum_{T_i^{(N)} \leq t} \left(\int_{T_{i-1}}^{T_i} (L_s^{\perp} - L_{i-1}^{\perp}) d\langle M, L^{\perp} \rangle_s \right) + \mathcal{O}_p(1) = \mathcal{O}_p(1)$$

converges to zero. The same holds true for every bounded $\mathcal{F}_t$-martingale orthogonal to M. Applying Theorem 1.6, we deduce that Proposition 3.3.2 holds true. □

Proposition 3.3.1 is a direct consequence of Proposition 3.3.2 since for $t = T$ the marginal distribution is simply a mixed normal distribution independent of $\mathcal{F}$. The stable convergence assures that the convergence also holds under the original probability measure and non-zero drift terms with the same asymptotic law. □

3.3.2 Error due to non-synchronicity

Proposition 3.3.4. *Let Assumptions 1, 2(a) and* (3.8b)-(3.8c) *from Assumption 3.1 be satisfied. The error* A_T^N *due to the lack of synchronicity converges stably in law to a centred mixed Gaussian distribution:*

$$\sqrt{\frac{N}{T}} A_T^N \overset{st}{\rightsquigarrow} \mathbf{N}\left(0, v_{A_T}\right) \ , \tag{3.14}$$

with asymptotic variance

$$v_{A_T} = \int_0^T F'(t) \left(\sigma_t^X \sigma_t^Y\right)^2 dt + \int_0^T 2H'(t) \left(\rho_t \sigma_t^X \sigma_t^Y\right)^2 dt \ . \tag{3.15}$$

Proof. First, we write the ith increments occurring as factors in the addends of the estimator (3.2) as the sum of the next-tick interpolation at T_i, the increments $\Delta X_{T_i} = X_{T_i} - X_{T_{i-1}}$ and $\Delta Y_{T_i} = Y_{T_i} - Y_{T_{i-1}}$, respectively, and the previous-tick interpolation at T_{i-1} and multiply out the addends.

$$\begin{aligned}\widehat{\langle X, Y \rangle}_T =& \sum_{i=1}^N (X_{g_i} - X_{T_i} + X_{T_i} - X_{T_{i-1}} + X_{T_{i-1}} - X_{l_i}) \left(Y_{\gamma_i} - Y_{T_i} + Y_{T_i} - Y_{T_{i-1}} + Y_{T_{i-1}} - Y_{\lambda_i}\right) \\ =& D_T^N + \sum_{i=1}^N \left((X_{g_i} - X_{T_i})\Delta Y_{T_i} + (Y_{\gamma_i} - Y_{T_i})\Delta X_{T_i} + (X_{T_{i-1}} - X_{l_i})\Delta Y_{T_i} \right. \\ & \left. + (Y_{T_{i-1}} - Y_{\lambda_i})\Delta X_{T_i} + (X_{g_i} - X_{T_i})(Y_{T_{i-1}} - Y_{\lambda_i}) + (Y_{\gamma_i} - Y_{T_i})(X_{T_{i-1}} - X_{l_i}) \right)\end{aligned}$$

The indicator functions in (3.6) have been dropped since the corresponding addends are zero if the indicator functions were zero. Since at least one of the next-tick interpolation errors is zero and as well one of the previous-tick interpolation errors, too, two addends, namely the products of next-tick interpolation errors and the product of previous-tick interpolation errors, equal zero. Thus, the error due to asynchronicity can be written as the sum of the remaining six terms (where at least another three equal zero in each addend). We conclude, that the error A_T^N can be expressed in the following way:

$$A_T^N = \sum_{i=1}^{N-1} ((X_{g_i} - X_{T_i})(Y_{T_i} - Y_{\lambda_i}) + (Y_{\gamma_i} - Y_{T_i})(X_{T_i} - X_{l_i})$$
$$(X_{T_{i+1}} - X_{T_i})(Y_{T_i} - Y_{\lambda_{i+1}}) + (X_{T_i} - X_{l_{i+1}})(Y_{T_{i+1}} - Y_{T_i})) + \mathcal{O}_p(1) \ .$$

In this equality an index shift has been applied to the partial sum of previous-tick interpolated errors multiplied with ΔX_{T_i} and ΔY_{T_i}, respectively, leading to the structure that in the ith addend the factors contain next- and previous-tick interpolated errors to the time T_i. The asymptotically negligible term emerges from the first and the last addend of the non-shifted original sum.
In the last illustration of A_T^N consecutive addends of the sum are uncorrelated in contrast to the non-shifted sum. The reason is that, if without loss of generality $\gamma_i = T_i$ holds, $(X_{g_i} - X_{T_i})\Delta Y_{T_i}$ and $(X_{T_i} - X_{l_{i+1}})\Delta Y_{T_{i+1}}$ have in general a non-zero correlation whereas $(X_{g_i} - X_{T_i})\Delta Y_{T_i}$ and $(X_{T_{i-1}} - X_{l_i})\Delta Y_{T_i}$ are uncorrelated. Furthermore, the fact that $\gamma_i = T_i \Rightarrow \lambda_{i+1} = T_i$ assures that the addends in the last illustration of A_T^N are uncorrelated.
As in the foregoing proof of Proposition 3.3.1, it is sufficient to prove the stable convergence result for the zero-drift case. We denote, as before, the corresponding transformed processes $L_t = \int_0^t \sigma_s^X dW_s^X$ and $M_t = \int_0^t \sigma_s^Y dW_s^Y$.
Consider the sum

$$\mathcal{A}_t^N = \sum_{T_{i+1}^{(N)} \le t} \Delta A_i^N := \sqrt{\frac{N}{T}} \sum_{T_{i+1}^{(N)} \le t} ((L_{g_i} - L_{T_i})(M_{T_i} - M_{\lambda_i}) + (M_{\gamma_i} - M_{T_i})(L_{T_i} - L_{l_i})$$
$$+(L_{T_i} - L_{l_{i+1}})(M_{T_{i+1}} - M_{T_i}) + (M_{T_i} - M_{\lambda_{i+1}})(L_{T_{i+1}} - L_{T_i})) \tag{3.16}$$

for fixed $0 \le t \le T$.

Proposition 3.3.5. *Assume the same conditions as in Proposition 3.3.4. For fixed $0 \le t \le T$ the transformed error due to non-synchronicity $\mathcal{A}_t^N$ is the endpoint of a discrete, centred, square-integrable martingale with respect to the filtration $\mathcal{F}_{i,N} := \mathcal{F}_{T_{i+1}^{(N)}}$. The process $\mathcal{A}_t^N$ converges as $N \to \infty$ stably in law:*

$$\mathcal{A}_t^N \overset{st}{\rightsquigarrow} \mathcal{A}_t = \int_0^t \sqrt{v_{\mathcal{A}_s}} dW_s^{\perp} \tag{3.17}$$

where $W^{\perp}$ is a Brownian motion independent of $\mathcal{F}$ and

$$v_{\mathcal{A}_s} = F'(s) \left(\sigma_s^X \sigma_s^Y\right)^2 + 2H'(s) \left(\rho_s \sigma_s^X \sigma_s^Y\right)^2 \ . \tag{3.18}$$

Proof. The expectation of the absolute value of the sum is bounded for all $t \in [0, T]$ and

ΔA_i^N, $i = 0, \ldots, N$ are $\mathcal{F}_{i,N} = \mathcal{F}_{T_{i+1}^{(N)}}$-measurable. Since

$$\begin{aligned}\mathbb{E}\left[\Delta A_i^N | \mathcal{F}_{i-1,N}\right] &= \mathbb{E}\left[\Delta A_i^N | \mathcal{F}_{T_i^{(N)}}\right] \\ &= \mathbb{E}\left[(L_{g_i} - L_{T_i})(M_{T_i} - M_{\lambda_i}) + (M_{\gamma_i} - M_{T_i})(L_{T_i} - L_{l_i})\right. \\ &\quad \left. + (L_{T_i} - L_{l_{i+1}})\Delta M_{T_{i+1}} + (M_{T_i} - M_{\lambda_{i+1}})\Delta L_{T_{i+1}} | \mathcal{F}_{T_i^{(N)}}\right] \\ &= \mathbb{E}\left[L_{g_i} - L_{T_i}\right](M_{T_i} - M_{\lambda_i}) + \mathbb{E}\left[M_{\gamma_i} - M_{T_i}\right](L_{T_i} - L_{l_i}) \\ &\quad + (L_{T_i} - L_{l_{i+1}})\mathbb{E}\left[\Delta M_{T_{i+1}}\right] + (M_{T_i} - M_{\lambda_{i+1}})\mathbb{E}\left[\Delta L_{T_{i+1}}\right] = 0\end{aligned}$$

for the conditional expectation of the increments holds, $\mathcal{A}_t^N$ is the endpoint of a $\mathcal{F}_{i,N}$-martingale.
The stable weak convergence to a limiting Brownian motion is proven with Corollary 1.3.5 to Jacod's Theorem 1.6.
First, we verify the conditional Lindeberg condition (C–LB) that is implied by the stronger conditional Lyapunov condition (C–LY). Therefore, we proof the following lemma:

Lemma 3.3.6. *The sum of the conditional fourth moments of the martingale increments A_i^N converges to zero in probability:*

$$\mathbb{E}\left[\sum_{T_{i+1}^{(N)} \le t} \left(\Delta A_i^N\right)^4 \Big| \mathcal{F}_{i-1,N}\right] = \mathcal{O}_p(1) \; .$$

Proof. Throughout the proof C denotes a generic constant that does not depend on N. We consider the addends of the fourth conditional moments consecutively. The sum of conditional fourth moments incorporates partial sums of the following types:

- fourth-order moments:

$$\frac{N^2}{T^2} \sum_{T_{i+1}^{(N)} \le t} \mathbb{E}\left[(L_{g_i} - L_{T_i})^4\right](M_{T_i} - M_{\lambda_i})^4 \; ,$$

- second-order moments:

$$\frac{N^2}{T^2} \sum_{T_{i+1}^{(N)} \le t} \mathbb{E}\left[(L_{g_i} - L_{T_i})^2 (\Delta M_{T_{i+1}})^2\right](L_{T_i} - L_{l_{i+1}})^2 (M_{T_i} - M_{\lambda_i})^2 \; ,$$

- third- and first-order moments:

$$\frac{N^2}{T^2} \sum_{T_{i+1}^{(N)} \le t} 4(M_{T_i} - M_{\lambda_i})^3 (L_{T_i} - L_{l_{i+1}})^3 \mathbb{E}\left[\Delta M_{T_{i+1}}(L_{g_i} - L_{T_i})\right] \; .$$

For the partial sum of the first type including fourth-order moments an application of the Burkholder-Davis-Gundy inequalities 1.1.4 yields

$$
\begin{aligned}
&\frac{N^2}{T^2}\sum_{T_{i+1}^{(N)}\le t}\mathbb{E}\left[(L_{g_i}-L_{T_i})^4\right](M_{T_i}-M_{\lambda_i})^4\\
\le&C\,\frac{N^2}{T^2}\sum_{T_{i+1}^{(N)}\le t}\mathbb{E}\left[\left(\int_{T_i}^{g_i}(\sigma_s^X)^2ds\right)^2\right](M_{T_i}-M_{\lambda_i})^4\\
\le&C\,\frac{N^2}{T^2}\sup_{s\in[0,T]}(\sigma_s^X)^2\sum_{T_{i+1}^{(N)}\le t}(M_{T_i}-M_{\lambda_i})^4(g_i-T_i)^2\le \mathcal{O}_p\left(N\delta_N^2\right)=\mathcal{o}_p(1)\;.
\end{aligned}
$$

The last inequality can be deduced by the result that the convergence $(N/(3T))\sum_i (\Delta M_{T_i})^4\to\int_0^t(\sigma_s^Y)^4ds$ holds almost surely as $N\to\infty$ for the so-called realized quarticity (Barndorff-Nielsen and Shephard [2002]) and that $(g_i-T_i)\le\delta_N$. Without the result about the convergence of the realized quarticity, the asymptotic order in probability can be derived by the convergence to zero of the expectation of the above sum and calculating the second moment that is bounded from above by a constant times $N^4\delta_N^7$.
For the partial sums incorporating second-order moments we obtain an upper bound by application of the Cauchy-Schwarz inequality and the Burkholder-Davis-Gundy inequalities:

$$
\begin{aligned}
&\frac{N^2}{T^2}\sum_{T_{i+1}^{(N)}\le t}6\,\mathbb{E}\left[(L_{g_i}-L_{T_i})^2(\Delta M_{T_{i+1}})^2\right](L_{T_i}-L_{l_{i+1}})^2(M_{T_i}-M_{\lambda_i})^2\\
\le&\frac{N^2}{T^2}\sum_{T_{i+1}^{(N)}\le t}6\sqrt{\mathbb{E}\left[(L_{g_i}-L_{T_i})^4\right]}\sqrt{\mathbb{E}\left[(\Delta M_{T_{i+1}})^4\right]}(L_{T_i}-L_{l_{i+1}})^2(M_{T_i}-M_{\lambda_i})^2\\
\le&C\,\frac{N^2}{T^2}\sum_{T_{i+1}^{(N)}\le t}6\left(\mathbb{E}\left[\left(\int_{T_i}^{g_i}(\sigma_s^X)^2ds\right)^2\right]\mathbb{E}\left[\left(\int_{T_i}^{T_{i+1}}(\sigma_s^Y)^2ds\right)^2\right]\right)^{\frac{1}{2}}\\
&\times(L_{T_i}-L_{l_{i+1}})^2(M_{T_i}-M_{\lambda_i})^2=\mathcal{o}_p(1)\;.
\end{aligned}
$$

The stochastic order follows, since the term has the expectation

$$
\begin{aligned}
&C\,\frac{N^2}{T^2}\sum_{T_{i+1}^{(N)}\le t}6\left(\mathbb{E}\left[\left(\int_{T_i}^{g_i}(\sigma_s^X)^2ds\right)^2\right]\mathbb{E}\left[\left(\int_{T_i}^{T_{i+1}}(\sigma_s^Y)^2ds\right)^2\right]\right)^{\frac{1}{2}}\\
&\quad\times\mathbb{E}\left[(L_{T_i}-L_{l_{i+1}})^2(M_{T_i}-M_{\lambda_i})^2\right]\\
&\le C\,\frac{N^2}{T^2}\sum_{T_{i+1}^{(N)}\le t}6\left(\mathbb{E}\left(\int_{T_i}^{g_i}(\sigma_s^X)^2ds\right)^2\mathbb{E}\left(\int_{T_i}^{T_{i+1}}(\sigma_s^Y)^2ds\right)^2\right)^{\frac{1}{2}}
\end{aligned}
$$

$$\times \mathbb{E}\left(\int_{l_{i+1}}^{T_i} (\sigma_s^X)^2 ds\right)^2 \left(\mathbb{E}\left(\int_{\lambda_i}^{T_i} (\sigma_s^Y)^2 ds\right)^2\right)^{\frac{1}{2}}$$
$$\leq C\, N^2 \delta_N^3 = \mathcal{O}(1) \ ,$$

where again the Cauchy-Schwarz and BDG-inequalities have been applied. The variance is bounded from above by a constant times $N^4\delta_N^7$ what can be shown by a similar calculation where thanks to the fact that $T_i = \gamma_i \Rightarrow \lambda_{i+1} = T_i$ the addends are uncorrelated and the variance of the sum equals the sum of variances.
We treat the third type of addends occurring in the sum of conditional fourth moments in the same way. Itô isometry yields

$$\frac{N^2}{T^2} \sum_{T_{i+1}^{(N)} \leq t} 4(M_{T_i} - M_{\lambda_i})^3 (L_{T_i} - L_{l_{i+1}})^3 \mathbb{E}\left[\Delta M_{T_{i+1}} (L_{g_i} - L_{T_i})\right]$$
$$= \frac{N^2}{T^2} \sum_{T_{i+1}^{(N)} \leq t} 4(M_{T_i} - M_{\lambda_i})^3 (L_{T_i} - L_{l_{i+1}})^3 \mathbb{E}\left[\int_{T_i}^{g_i} \rho_s \sigma_s^X \sigma_s^Y ds\right] \ .$$

This term has expectation

$$\frac{N^2}{T^2} \sum_{T_{i+1}^{(N)} \leq t} 4\mathbb{E}\left[(M_{T_i} - M_{\lambda_i})^3 (L_{T_i} - L_{l_{i+1}})^3\right] \mathbb{E}\left[\int_{T_i}^{g_i} \rho_s \sigma_s^X \sigma_s^Y ds\right]$$
$$\leq \frac{N^2}{T^2} \sum_{T_{i+1}^{(N)} \leq t} 4\sqrt{\mathbb{E}\left[(M_{T_i} - M_{\lambda_i})^6\right] \mathbb{E}\left[(L_{T_i} - L_{l_{i+1}})^6\right]} \mathbb{E}\left[\int_{T_i}^{g_i} \rho_s \sigma_s^X \sigma_s^Y ds\right]$$
$$\leq C\, \frac{N^2}{T^2} \sum_{T_{i+1}^{(N)} \leq t} 4\left(\mathbb{E}\left[\left(\int_{\lambda_i}^{T_i} (\sigma_s^Y)^2 ds\right)^3\right] \mathbb{E}\left[\left(\int_{l_{i+1}}^{T_i} (\sigma_s^X)^2 ds\right)^3\right]\right)^{1/2} \mathbb{E}\left[\int_{T_i}^{g_i} \rho_s \sigma_s^X \sigma_s^Y ds\right]$$
$$\leq C N^2 \delta_N^3 = \mathcal{O}(1) \ ,$$

and an analogous calculation as before yields that the variance is of asymptotic order $N^4\delta_N^7$.
Since all addends in the sum of conditional fourth moments are of one of the three above considered types, the sum converges to zero in probability and hence the conditional Lyapunov condition of Lemma 3.3.6 holds true. □

Next, we consider the sum of conditional variances of the increments of the discrete martingale.

Lemma 3.3.7.

$$\mathbb{E}\left[\sum_{T_{i+1}^{(N)}\le t}\left(\Delta A_i^N\right)^2\Big|\mathcal{F}_{T_i^{(N)}}\right]\xrightarrow{p}\int_0^t F'(s)\left(\sigma_s^X\sigma_s^Y\right)^2ds+\int_0^t 2H'(s)\left(\rho_s\sigma_s^X\sigma_s^Y\right)^2ds\ . \tag{3.19}$$

Proof.

$$\begin{aligned}
&\mathbb{E}\left[\sum_{T_{i+1}^{(N)}\le t}\left(\Delta A_i^N\right)^2\Big|\mathcal{F}_{T_i^{(N)}}\right]=\frac{N}{T}\sum_{T_{i+1}^{(N)}\le t}\mathbb{E}\Big[(L_{g_i}-L_{T_i})^2(M_{T_i}-M_{\lambda_i})^2\\
&\times(M_{\gamma_i}-M_{T_i})^2(L_{T_i}-L_{l_i})^2+(L_{T_i}-L_{l_{i+1}})^2(\Delta M_{T_{i+1}})^2\\
&+(M_{T_i}-M_{\lambda_{i+1}})^2(\Delta L_{T_{i+1}})^2+2(L_{g_i}-L_{T_i})(M_{T_i}-M_{\lambda_i})(L_{T_i}-L_{l_{i+1}})\Delta M_{T_{i+1}}\\
&+2(M_{\gamma_i}-M_{T_i})(L_{T_i}-L_{l_i})(M_{T_i}-M_{l_{i+1}})\Delta L_{T_{i+1}}\Big|\mathcal{F}_{T_i^{(N)}}\Big]\\
&\underset{\text{(Itô isometry)}}{=}\frac{N}{T}\sum_{T_{i+1}^{(N)}\le t}\Bigg(\mathbb{E}\left[\int_{T_i}^{g_i}(\sigma_s^X)^2ds\right](M_{T_i}-M_{\lambda_i})^2+\mathbb{E}\left[\int_{T_i}^{\gamma_i}(\sigma_s^Y)^2ds\right](L_{T_i}-L_{l_i})^2\\
&+(L_{T_i}-L_{l_{i+1}})^2\mathbb{E}\left[\int_{T_i}^{T_{i+1}}(\sigma_s^Y)^2ds\right]+(M_{T_i}-M_{\lambda_{i+1}})^2\mathbb{E}\left[\int_{T_i}^{T_{i+1}}(\sigma_s^X)^2ds\right]\\
&+2(M_{T_i}-M_{\lambda_i})(L_{T_i}-L_{l_{i+1}})\mathbb{E}\left[\int_{T_i}^{g_i}\rho_s\sigma_s^X\sigma_s^Y\,ds\right]\\
&+2(L_{T_i}-L_{l_i})(M_{T_i}-M_{l_{i+1}})\mathbb{E}\left[\int_{T_i}^{\gamma_i}\rho_s\sigma_s^X\sigma_s^Y\,ds\right]\Bigg)\\
&\underset{\text{(Lemma 3.3.8)}}{=}\frac{N}{T}\sum_{T_{i+1}^{(N)}\le t}\Bigg(\mathbb{E}\left[\int_{T_i}^{g_i}(\sigma_s^X)^2ds\right]\int_{\lambda_i}^{T_i}(\sigma_s^Y)^2ds+\mathbb{E}\left[\int_{T_i}^{\gamma_i}(\sigma_s^Y)^2ds\right]\int_{l_i}^{T_i}(\sigma_s^X)^2ds\\
&+\int_{l_{i+1}}^{T_i}(\sigma_s^X)^2ds\,\mathbb{E}\left[\int_{T_i}^{T_{i+1}}(\sigma_s^Y)^2ds\right]+\int_{\lambda_{i+1}}^{T_i}(\sigma_s^Y)^2ds\,\mathbb{E}\left[\int_{T_i}^{T_{i+1}}(\sigma_s^X)^2ds\right]\\
&+\int_{l_{i+1}}^{T_i}\rho_s\sigma_s^X\sigma_s^Y\,ds\,\mathbb{E}\left[\int_{T_i}^{g_i}\rho_s\sigma_s^X\sigma_s^Y\,ds\right]+\int_{l_{i+1}}^{T_i}\rho_s\sigma_s^X\sigma_s^Y\,ds\,\mathbb{E}\left[\int_{T_i}^{\gamma_i}\rho_s\sigma_s^X\sigma_s^Y\,ds\right]\Bigg)+\mathcal{O}_p(1)\\
&\underset{\text{(Lemma 3.3.8)}}{=}\frac{N}{T}\sum_{T_{i+1}^{(N)}\le t}\Bigg(\int_{T_i}^{g_i}(\sigma_s^X)^2ds\int_{\lambda_i}^{T_i}(\sigma_s^Y)^2ds+\int_{T_i}^{\gamma_i}(\sigma_s^Y)^2ds\int_{l_i}^{T_i}(\sigma_s^X)^2ds\\
&+\int_{l_{i+1}}^{T_i}(\sigma_s^X)^2ds\int_{T_i}^{T_{i+1}}(\sigma_s^Y)^2ds+\int_{\lambda_{i+1}}^{T_i}(\sigma_s^Y)^2ds\int_{T_i}^{T_{i+1}}(\sigma_s^X)^2ds\\
&+2\int_{l_{i+1}}^{T_i}\rho_s\sigma_s^X\sigma_s^Y\,ds\int_{T_i}^{g_i}\rho_s\sigma_s^X\sigma_s^Y\,ds+2\int_{l_{i+1}}^{T_i}\rho_s\sigma_s^X\sigma_s^Y\,ds\int_{T_i}^{\gamma_i}\rho_s\sigma_s^X\sigma_s^Y\,ds\Bigg)+\mathcal{O}_p(1)
\end{aligned}$$

$$\underset{(\text{Lemma 3.3.9})}{=}\frac{N}{T}\sum_{T_{i+1}^{(N)}\leq t}\Big((\sigma_{T_i}^X\sigma_{T_i}^Y)^2\left((T_i-\lambda_i)(g_i-T_i)+(\gamma_i-T_i)(T_i-l_i)+(T_i-l_{i+1})\Delta T_{i+1}\right.$$

$$\left.+(T_i-\lambda_{i+1})\Delta T_{i+1}\right)+(\rho_{T_i}\sigma_{T_i}^X\sigma_{T_i}^Y)^2(2(T_i-l_{i+1})(g_i-T_i)+2(T_i-\lambda_{i+1})(\gamma_i-T_i))\Big)+\mathcal{O}_p(1)$$

$$=\sum_{T_{i+1}^{(N)}\leq t}\frac{F(T_{i+1})-F(T_i)}{T_{i+1}-T_i}(\sigma_{T_i}^X\sigma_{T_i}^Y)^2\Delta T_{i+1}+2\frac{H(T_{i+1})-H(T_i)}{T_{i+1}-T_i}(\rho_{T_i}\sigma_{T_i}^X\sigma_{T_i}^Y)^2\Delta T_{i+1}+\mathcal{O}_p(1)$$

$$\xrightarrow{p}\int_0^t F'(s)\left(\sigma_s^X\sigma_s^Y\right)^2ds+\int_0^t 2H'(s)\left(\rho_s\sigma_s^X\sigma_s^Y\right)^2ds\ .$$

In the first equality Itô isometry has been used. The proofs of the following three equalities are postponed to the next two lemmas. In the last step we have involved Definition 3.2.3. The Riemann sum converges on the Assumption 3.1 (in particular (3.8b) and (3.8c)) in probability as $N\to\infty$ to the expression $\int_0^t v_{\mathcal{A}_s}ds$ with $v_{\mathcal{A}_s}$ given in Proposition 3.3.5.

Lemma 3.3.8. *On the same assumptions as before, the following equations hold true:*

$$\frac{N}{T}\sum_{T_{i+1}^{(N)}\leq t}\left((M_{T_i}-M_{\lambda_i})^2-\int_{\lambda_i}^{T_i}(\sigma_s^Y)^2ds\right)\mathbb{E}\left[\int_{T_i}^{g_i}(\sigma_s^X)^2ds\right]=\mathcal{O}_p(1)\ ,$$

$$\frac{N}{T}\sum_{T_{i+1}^{(N)}\leq t}\int_{\lambda_i}^{T_i}(\sigma_s^Y)^2ds\left(\mathbb{E}\left[\int_{T_i}^{g_i}(\sigma_s^X)^2ds\right]-\int_{T_i}^{g_i}(\sigma_s^X)^2ds\right)=\mathcal{O}_p(1)\ ,$$

$$\frac{N}{T}\sum_{T_{i+1}^{(N)}\leq t}\left((L_{T_i}-L_{l_i})^2-\int_{l_i}^{T_i}(\sigma_s^X)^2ds\right)\mathbb{E}\left[\int_{T_i}^{\gamma_i}(\sigma_s^Y)^2ds\right]=\mathcal{O}_p(1)\ ,$$

$$\frac{N}{T}\sum_{T_{i+1}^{(N)}\leq t}\int_{l_i}^{T_i}(\sigma_s^X)^2ds\left(\mathbb{E}\left[\int_{T_i}^{\gamma_i}(\sigma_s^Y)^2ds\right]-\int_{T_i}^{\gamma_i}(\sigma_s^Y)^2ds\right)=\mathcal{O}_p(1)\ ,$$

$$\frac{N}{T}\sum_{T_{i+1}^{(N)}\leq t}\left((M_{T_i}-M_{\lambda_{i+1}})^2-\int_{\lambda_{i+1}}^{T_i}(\sigma_s^Y)^2ds\right)\mathbb{E}\left[\int_{T_i}^{T_{i+1}}(\sigma_s^X)^2ds\right]=\mathcal{O}_p(1)\ ,$$

$$\frac{N}{T}\sum_{T_{i+1}^{(N)}\leq t}\int_{\lambda_{i+1}}^{T_i}(\sigma_s^Y)^2ds\left(\mathbb{E}\left[\int_{T_i}^{T_{i+1}}(\sigma_s^X)^2ds\right]-\int_{T_i}^{T_{i+1}}(\sigma_s^X)^2ds\right)=\mathcal{O}_p(1)\ ,$$

$$\frac{N}{T}\sum_{T_{i+1}^{(N)}\le t}\left((L_{T_i}-L_{l_{i+1}})^2-\int_{l_{i+1}}^{T_i}(\sigma_s^X)^2ds\right)\mathbb{E}\left[\int_{T_i}^{T_{i+1}}(\sigma_s^Y)^2ds\right]=\mathcal{o}_p(1)\ ,$$

$$\frac{N}{T}\sum_{T_{i+1}^{(N)}\le t}\int_{l_{i+1}}^{T_i}(\sigma_s^X)^2ds\left(\mathbb{E}\left[\int_{T_i}^{T_{i+1}}(\sigma_s^Y)^2ds\right]-\int_{T_i}^{T_{i+1}}(\sigma_s^Y)^2ds\right)=\mathcal{o}_p(1)\ ,$$

$$\frac{N}{T}\sum_{T_{i+1}^{(N)}\le t}\left((M_{T_i}-M_{\lambda_i})(L_{T_i}-L_{l_{i+1}})-\int_{l_{i+1}}^{T_i}\rho_s\sigma_s^X\sigma_s^Y ds\right)\mathbb{E}\left[\int_{T_i}^{g_i}\rho_s\sigma_s^X\sigma_s^Y ds\right]=\mathcal{o}_p(1)\ ,$$

$$\frac{N}{T}\sum_{T_{i+1}^{(N)}\le t}\int_{l_{i+1}}^{T_i}\rho_s\sigma_s^X\sigma_s^Y ds\left(\mathbb{E}\left[\int_{T_i}^{g_i}\rho_s\sigma_s^X\sigma_s^Y ds\right]-\int_{T_i}^{g_i}\rho_s\sigma_s^X\sigma_s^Y ds\right)=\mathcal{o}_p(1)\ ,$$

$$\frac{N}{T}\sum_{T_{i+1}^{(N)}\le t}\left((L_{T_i}-L_{l_i})(M_{T_i}-L_{\lambda_{i+1}})-\int_{\lambda_{i+1}}^{T_i}\rho_s\sigma_s^X\sigma_s^Y ds\right)\mathbb{E}\left[\int_{T_i}^{\gamma_i}\rho_s\sigma_s^X\sigma_s^Y ds\right]=\mathcal{o}_p(1)\ ,$$

$$\frac{N}{T}\sum_{T_{i+1}^{(N)}\le t}\int_{\lambda_{i+1}}^{T_i}\rho_s\sigma_s^X\sigma_s^Y ds\left(\mathbb{E}\left[\int_{T_i}^{\gamma_i}\rho_s\sigma_s^X\sigma_s^Y ds\right]-\int_{T_i}^{\gamma_i}\rho_s\sigma_s^X\sigma_s^Y ds\right)=\mathcal{o}_p(1)\ .$$

Proof. We restrict ourselves to the proof of the first two equalities, since all other terms can shown to converge to zero in probability in an analogous way. The left-hand side of the first equality has an expectation equal to zero what can be concluded directly by Itô isometry:

$$\mathbb{E}\left[\frac{N}{T}\sum_{T_{i+1}^{(N)}\le t}\left((M_{T_i}-M_{\lambda_i})^2-\int_{\lambda_i}^{T_i}(\sigma_s^Y)^2ds\right)\mathbb{E}\left[\int_{T_i}^{g_i}(\sigma_s^X)^2ds\right]\right]=0\ .$$

In order to derive the stochastic order of the term, consider the second moment:

$$\mathbb{E}\left[\left(\frac{N}{T}\sum_{T_{i+1}^{(N)}\le t}\left((M_{T_i}-M_{\lambda_i})^2-\int_{\lambda_i}^{T_i}(\sigma_s^Y)^2ds\right)\mathbb{E}\left[\int_{T_i}^{g_i}(\sigma_s^X)^2ds\right]\right)^2\right]$$
$$=\frac{N^2}{T^2}\sum_{T_{i+1}^{(N)}\le t}\mathbb{E}\left[(M_{T_i}-M_{\lambda_i})^4-2(M_{T_i}-M_{\lambda_i})^2\int_{\lambda_i}^{T_i}(\sigma_s^Y)^2ds+\left(\int_{\lambda_i}^{T_i}(\sigma_s^Y)^2ds\right)^2\right]$$

$$\times \left(\mathbb{E}\left[\int_{T_i}^{g_i} (\sigma_s^X)^2 ds\right]\right)^2 = \mathcal{O}(1) \ ,$$

where the asymptotic order is deduced by Itô isometry and the fourth moment of Brownian increments (or application of the BDG inequalities). Since the error induced by this term in the approximation of the conditional variance before is centred and has a variance converging to zero as $N \to \infty$, the error is asymptotically negligible in the sense that it converges to zero in probability.
In the second equality we consider the error when the expected increment of the quadratic variation of X over the next-tick interpolated time interval is substituted by the integral itself. We proceed as before for the first approximation. Since

$$\mathbb{E}\left[\frac{N}{T}\sum_{T_{i+1}^{(N)}\le t}\int_{\lambda_i}^{T_i}(\sigma_s^Y)^2 ds\left(\mathbb{E}\left[\int_{T_i}^{g_i}(\sigma_s^X)^2 ds\right]-\int_{T_i}^{g_i}(\sigma_s^X)^2 ds\right)\right]=0$$

and the second moment

$$\mathbb{E}\left[\left(\frac{N}{T}\sum_{T_{i+1}^{(N)}\le t}\int_{\lambda_i}^{T_i}(\sigma_s^Y)^2 ds\left(\mathbb{E}\left[\int_{T_i}^{g_i}(\sigma_s^X)^2 ds\right]-\int_{T_i}^{g_i}(\sigma_s^X)^2 ds\right)\right)^2\right]$$
$$=\frac{N^2}{T^2}\sum_{T_{i+1}^{(N)}\le t}\mathbb{V}\mathrm{ar}\left(\int_{\lambda_i}^{T_i}(\sigma_s^Y)^2 ds\right)\mathbb{E}\left[\left(\int_{\lambda_i}^{T_i}(\sigma_s^Y)^2 ds\right)^2\right]=\mathcal{O}(1)$$

is bounded from above by a constant times $N^2\delta_N^3$ again, the approximation error is asymptotically negligible. The fact that $\gamma_i = T_i \Rightarrow \lambda_{i+1} = T_i$ has been used that guarantees that the addends of the sum are uncorrelated. ☐

Lemma 3.3.8 has been applied in the second and third equality in the sum of the conditional variances and the following Lemma 3.3.9 will complete the proof of Lemma 3.3.7.

Lemma 3.3.9. *On the same assumptions as before, the following equation holds true*

$$\frac{N}{T}\sum_{T_{i+1}^{(N)}\le t}\left(\int_{T_i}^{g_i}(\sigma_s^X)^2 ds\int_{\lambda_i}^{T_i}(\sigma_s^Y)^2 ds-(\sigma_{T_i}^X\sigma_{T_i}^Y)^2(T_i-\lambda_i)(g_i-T_i)\right)=\mathcal{O}_p(1)$$

and analogously the errors in the five other addends converge to zero in probability when replacing the product of increments of quadratic (co-)variations by the values of $\rho_{T_i}, \sigma_{T_i}^X, \sigma_{T_i}^Y$ *multiplied with the corresponding times increments.*

Proof. We prove the equality explicitly given in the lemma. The five remaining terms can be handled with the same strategy. By an application of the mean value theorem, elementary algebra and the triangle inequality for the absolute values, we deduce

$$
\begin{aligned}
&\frac{N}{T}\sum_{T_{i+1}^{(N)}\le t}\left(\int_{T_i}^{g_i}(\sigma_s^X)^2ds\int_{\lambda_i}^{T_i}(\sigma_s^Y)^2ds-(\sigma_{T_i}^X\sigma_{T_i}^Y)^2(T_i-\lambda_i)(g_i-T_i)\right)\\
&=\frac{N}{T}\sum_{T_{i+1}^{(N)}\le t}\left((\sigma_{\varsigma_i}^X\sigma_{\xi_i}^Y)^2-(\sigma_{T_i}^X\sigma_{T_i}^Y)^2\right)(T_i-\lambda_i)(g_i-T_i)\\
&\le\frac{N}{T}\sum_{T_{i+1}^{(N)}\le t}\left|(\sigma_{\varsigma_i}^X\sigma_{\xi_i}^Y)^2-(\sigma_{T_i}^X\sigma_{T_i}^Y)^2\right|(T_i-\lambda_i)(g_i-T_i)\\
&\le\frac{N}{T}\sum_{T_{i+1}^{(N)}\le t}\left(\left((\sigma_{\xi_i}^Y)^2\left|(\sigma_{\varsigma_i}^X)^2-(\sigma_{T_i}^X)^2\right|+(\sigma_{T_i}^X)^2\left|(\sigma_{\xi_i}^Y)^2-(\sigma_{T_i}^Y)^2\right|\right)(T_i-\lambda_i)(g_i-T_i)\right)\\
&\le C\,\frac{N}{T}\sum_{T_{i+1}^{(N)}\le t}\left(\sup_{s\in[\lambda_i,T_i]}\left|(\sigma_s^X)^2-(\sigma_{T_i}^X)^2\right|+\sup_{s\in[T_i,g_i]}\left|(\sigma_s^Y)^2-(\sigma_{T_i}^Y)^2\right|\right)(T_i-\lambda_i)(g_i-T_i)\\
&=\mathcal{O}_p(1)
\end{aligned}
$$

what is assured by the conditions of Assumption 1 on the volatility processes. □

□

For the stable convergence in Proposition 3.3.5 it remains to show that the discrete covariations of $\mathcal{A}_t^N$ with the $\mathcal{F}$-generating underlying martingales L_t and M_t converge to zero in probability.

Lemma 3.3.10.

$$
\sum_{T_{i+1}^{(N)}\le t}\mathbb{E}\left[\Delta A_i^N\Delta L_{T_{i+1}^{(N)}}\Big|\mathcal{F}_{T_i^{(N)}}\right]\stackrel{p}{\to}0\ ,
$$

$$
\sum_{T_{i+1}^{(N)}\le t}\mathbb{E}\left[\Delta A_i^N\Delta M_{T_{i+1}^{(N)}}\Big|\mathcal{F}_{T_i^{(N)}}\right]\stackrel{p}{\to}0\ .
$$

Proof. Both relations are proven similarly and we leave out the second one. The left-hand side of the first equation equals

$$
\begin{aligned}
&\sqrt{\frac{N}{T}}\sum_{T_{i+1}^{(N)}\le t}\mathbb{E}\left[\Delta L_{T_{i+1}}\left((L_{g_i}-L_{T_i})(M_{T_i}-M_{\lambda_i})+(M_{\gamma_i}-M_{T_i})(L_{T_i}-L_{l_i})\right.\right.\\
&\qquad\left.\left.+(L_{T_i}-L_{l_{i+1}})(M_{T_{i+1}}-M_{T_i})+(M_{T_i}-M_{\lambda_{i+1}})(L_{T_{i+1}}-L_{T_i})\right)\Big|\mathcal{F}_{T_i{}^{(N)}}\right]\\
&=\sqrt{\frac{N}{T}}\sum_{T_{i+1}^{(N)}\le t}\left(\mathbb{E}\left[\int_{T_i}^{g_i}(\sigma_s^X)^2ds\right](M_{T_i}-M_{\lambda_i})+\mathbb{E}\left[\int_{T_i}^{\gamma_i}\rho_s\sigma_s^X\sigma_s^Y ds\right](L_{T_i}-L_{l_i})\right.
\end{aligned}
$$

$$+(L_{T_i}-L_{l_{i+1}})\mathbb{E}\left[\int_{T_i}^{T_{i+1}}\rho_s\sigma_s^X\sigma_s^Y ds\right]+(M_{T_i}-M_{\lambda_{i+1}})\mathbb{E}\left[\int_{T_i}^{T_{i+1}}(\sigma_s^X)^2 ds\right]\Bigg)=:\Gamma\ .$$

Γ is centred and calculating the variance using Itô isometry yields

$$\frac{N}{T}\sum_{T_{i+1}^{(N)}\le t}\Bigg(\left(\mathbb{E}\left[\int_{T_i}^{g_i}(\sigma_s^X)^2ds\right]\right)^2\mathbb{E}\left[\int_{\lambda_i}^{T_i}(\sigma_s^Y)^2ds\right]+\left(\mathbb{E}\left[\int_{T_i}^{\gamma_i}\rho_s\sigma_s^X\sigma_s^Y ds\right]\right)^2\mathbb{E}\left[\int_{l_i}^{T_i}(\sigma_s^X)^2ds\right]$$
$$+\mathbb{E}\left[\int_{l_{i+1}}^{T_i}(\sigma_s^X)^2ds\right]\left(\mathbb{E}\left[\int_{T_i}^{T_{i+1}}\rho_s\sigma_s^X\sigma_s^Y ds\right]\right)^2+\mathbb{E}\left[\int_{\lambda_{i+1}}^{T_i}(\sigma_s^Y)^2ds\right]\left(\mathbb{E}\left[\int_{T_i}^{T_{i+1}}(\sigma_s^X)^2ds\right]\right)^2$$
$$+2\mathbb{E}\left[\int_{l_{i+1}}^{T_i}\rho_s\sigma_s^X\sigma_s^Y ds\right]\mathbb{E}\left[\int_{T_i}^{g_i}(\sigma_s^X)^2ds\right]\mathbb{E}\left[\int_{T_i}^{T_{i+1}}\rho_s\sigma_s^X\sigma_s^Y ds\right]$$
$$+2\mathbb{E}\left[\int_{\lambda_{i+1}}^{T_i}\rho_s\sigma_s^Y\sigma_s^Y ds\right]\mathbb{E}\left[\int_{T_i}^{T_{i+1}}(\sigma_s^X)^2ds\right]\mathbb{E}\left[\int_{T_i}^{\gamma_i}\rho_s\sigma_s^X\sigma_s^Y ds\right]\Bigg)$$
$$\le CN\delta_N^2=\mathcal{O}(1)\ .$$

Once more we can conclude that the addends are uncorrelated since $\gamma_i = T_i \Rightarrow \lambda_{i+1} = T_i$ and $g_i = T_i \Rightarrow l_{i+1} = T_i$, respectively. □

We are left to verify the last condition of the discrete-time version to Jacod's Theorem 1.6 in Corollary 1.3.5. It suffices to prove that the discrete covariation of the martingale with every bounded $\mathcal{F}_t$-adapted martingale, that is orthogonal to L_t and M_t converges to zero in probability. From this result, we are able to conclude the stability of the convergence and the asymptotic independence of the limiting Brownian motion is established that is defined on an orthogonal extension of the original underlying probability space. In the next lemma, we can even prove the stronger result, that the discrete covariation of our considered martingale with every bounded $\mathcal{F}_t$-martingale that is orthogonal to L_t or M_t, converges to zero in probability. Hence, this lemma will complete the proof of Proposition 3.3.5.

Lemma 3.3.11. *Assume that $L_t^\perp$ and $M_t^\perp$ are bounded $\mathcal{F}_t$-martingales, with $\langle L, L^\perp\rangle \equiv 0$ and $\langle M, M^\perp\rangle \equiv 0$, respectively. It holds true that*

$$\sum_{T_{i+1}^{(N)}\le t}\mathbb{E}\left[\Delta A_i^N\Delta L^\perp_{T_{i+1}^{(N)}}\Big|\mathcal{F}_{T_i^{(N)}}\right]\xrightarrow{p}0\ ,$$
$$\sum_{T_{i+1}^{(N)}\le t}\mathbb{E}\left[\Delta A_i^N\Delta M^\perp_{T_{i+1}^{(N)}}\Big|\mathcal{F}_{T_i^{(N)}}\right]\xrightarrow{p}0\ .$$

Proof. As in the preceding lemma, we only prove the first part of the result. The left-hand side of the first equation equals

$$\sqrt{\frac{N}{T}}\sum_{T_{i+1}^{(N)}\le t}\mathbb{E}\Big[\Delta L^{\perp}_{T_{i+1}}\left((L_{g_i}-L_{T_i})(M_{T_i}-M_{\lambda_i})+(M_{\gamma_i}-M_{T_i})(L_{T_i}-L_{l_i})\right.$$
$$\left.+(L_{T_i}-L_{l_{i+1}})(M_{T_{i+1}}-M_{T_i})+(M_{T_i}-M_{\lambda_{i+1}})(L_{T_{i+1}}-L_{T_i})\right)\Big|\mathcal{F}_{T_i(N)}\Big]$$
$$=\sqrt{\frac{N}{T}}\sum_{T_{i+1}^{(N)}\le t}\Bigg(\mathbb{E}\left[\int_{T_i}^{g_i}d\langle L,L^{\perp}\rangle_s\right](M_{T_i}-M_{\lambda_i})+\mathbb{E}\left[\int_{T_i}^{\gamma_i}d\langle M,L^{\perp}\rangle_s\right](L_{T_i}-L_{l_i})$$
$$+(L_{T_i}-L_{l_{i+1}})\mathbb{E}\left[\int_{T_i}^{T_{i+1}}d\langle M,L^{\perp}\rangle_s\right]+(M_{T_i}-M_{\lambda_{i+1}})\mathbb{E}\left[\int_{T_i}^{T_{i+1}}d\langle L,L^{\perp}\rangle_s\right]\Bigg)$$
$$=\sqrt{\frac{N}{T}}\sum_{T_{i+1}^{(N)}\le t}\mathbb{E}\left[\int_{T_i}^{T_{i+1}}d\langle M,L^{\perp}\rangle_s\right](L_{T_i}-L_{l_{i+1}})+\mathbb{E}\left[\int_{T_i}^{\gamma_i}d\langle M,L^{\perp}\rangle_s\right](L_{T_i}-L_{l_i})\ .$$

This term is centred and the has the variance

$$\frac{N}{T}\sum_{T_{i+1}^{(N)}\le t}\mathbb{E}\left[\left(\int_{T_i}^{T_{i+1}}d\langle M,L^{\perp}\rangle_s\right)^2\right]\mathbb{E}\left[(L_{T_i}-L_{l_{i+1}})^2\right]+\mathbb{E}\left[\left(\int_{T_i}^{\gamma_i}d\langle M,L^{\perp}\rangle_s\right)^2\right]$$
$$\times\mathbb{E}\left[(L_{T_i}-L_{l_i})^2\right]+\mathbb{E}\left[(L_{T_i}-L_{L_{i+1}})^2\right]\mathbb{E}\left[\int_{T_i}^{T_{i+1}}d\langle M,L^{\perp}\rangle_s\right]\mathbb{E}\left[\int_{T_i}^{\gamma_i}d\langle M,L^{\perp}\rangle_s\right]$$
$$\underset{\text{Itô isometry}}{=}\frac{N}{T}\sum_{T_{i+1}^{(N)}\le t}\mathbb{E}\left[\left(\int_{T_i}^{T_{i+1}}d\langle M,L^{\perp}\rangle_s\right)^2\right]\mathbb{E}\left[\int_{l_{i+1}}^{T_i}(\sigma_s^X)^2ds\right]$$
$$+\mathbb{E}\left[\left(\int_{T_i}^{\gamma_i}d\langle M,L^{\perp}\rangle_s\right)^2\right]\mathbb{E}\left[\int_{l_i}^{T_i}(\sigma_s^X)^2ds\right]$$
$$+\mathbb{E}\left[\int_{l_{i+1}}^{T_i}(\sigma_s^X)^2ds\right]\mathbb{E}\left[\int_{T_i}^{T_{i+1}}d\langle M,L^{\perp}\rangle_s\right]\mathbb{E}\left[\int_{T_i}^{\gamma_i}d\langle M,L^{\perp}\rangle_s\right]=\mathcal{O}(1)\ .$$

Thus, the covariations converge to zero in probability. □

The Lemma completes the proof of Proposition 3.3.5. □

The mixed normal limit in Proposition 3.3.4 can be obtained as the marginal distribution of $\mathcal{A}_T^N$ setting $t=T$. Proposition 3.3.4 is implied by the stronger result of Proposition 3.3.5 and hence the stable convergence of the error due to the lack of synchronicity has been proved. □

Proposition 3.3.4 for the error of the approximation by the discretization error of the closest synchronous approximation (3.6) and the stable limit theorem for this synchronous discretization error (3.5) given in Proposition 3.3.1 suffice to imply Theorem 3.1. The multivariate stable convergence Theorem 1.2.4 applies to the vector of the two uncorrelated terms. Since the covariations converge to zero the stable convergence to the mixed Gaussian limit with the sum of the two asymptotic variances is concluded.

3.4 The synchronized realized covariance under the influence of microstructure noise

In this section we concisely show that the synchronized realized covariance (3.2), namely the Hayashi-Yoshida estimator that has been presented in Section 3.1, loses consistency if the observations of X and Y are contaminated with noise. This is not surprising since the estimator is only adapted to non-synchronicity and besides that keeps to the structure of a realized covariance estimator evaluated with all available observations. We detect this in the following proposition.

Proposition 3.4.1. *On the Assumptions 2(a) and that processes*

$$\tilde{X}_{t_i} = X_{t_i} + \epsilon^X_{t_i}\,, i \in \{0, \ldots, n\}$$

$$\textit{and } \tilde{Y}_{\tau_j} = Y_{\tau_j} + \epsilon^Y_{\tau_j}\,, j \in \{0, \ldots, m\}$$

are observed, where X and Y fulfill Assumption 1 and the observation errors are i. i. d. centred, independent of each other and of X and Y, with finite variances η_X^2, η_Y^2, it holds true that

$$\mathbb{V}\mathrm{ar}_{X,Y}\left(\widehat{\langle X, Y\rangle}_T^{(HY)}\right) = 4N\eta_X^2\eta_Y^2 + \mathcal{O}_p(1)\ , \tag{3.20}$$

where $\mathbb{V}\mathrm{ar}_{X,Y}(\,\cdot\,)$ denotes the variance of the error due to noise conditional on the paths of X and Y.

Proof. The conditional variance can be simplified to

$$\mathbb{V}\mathrm{ar}_{X,Y}\left(\widehat{\langle X, Y\rangle}_T^{(HY)}\right) = \mathbb{E}\left[\sum_{i=0}^N \left(\epsilon^X_{g_i} - \epsilon^X_{l_i}\right)\left(\epsilon^Y_{\gamma_i} - \epsilon^Y_{\lambda_i}\right)\right]^2 + \mathcal{O}_p(1) =$$
$$\mathbb{E}\left[\sum_{i=0}^N \left(\epsilon^X_{g_i} - \epsilon^X_{l_i}\right)^2\left(\epsilon^Y_{\gamma_i} - \epsilon^Y_{\lambda_i}\right)^2\right] + \mathcal{O}_p(1) = 4N\eta_X^2\eta_Y^2 + \mathcal{O}_p(1)\,.$$

The variances of $\sum_i \left(\epsilon^X_{g_i} - \epsilon^X_{l_i}\right)(Y_{\gamma_i} - Y_{\lambda_i})$ and the second sum including increments of X and ϵ^Y lead to the term of order 1 in probability. The cross terms in the remaining second moment have an expectation equal to zero although consecutive sets $\mathcal{H}^i$ and $\mathcal{H}^{i+1}$ (or $\mathcal{G}^i$ and $\mathcal{G}^{i+1}$) are not generally disjoint. Nevertheless, our synchronization method was defined such that if the intersection of $\mathcal{H}^i$ and $\mathcal{H}^{i+1}$ is non-empty, $\mathcal{G}^i \cap \mathcal{G}^{i+1} = \emptyset$ holds. Assumption 3 yields that each addend has expectation $2\eta_X^2 \cdot 2\eta_Y^2$. □

The estimator will still be asymptotically unbiased when the noise processes are mutually independent but the variance explodes. We conclude that in the presence of noise the Hayashi-Yoshida estimator cannot yield an adequate estimation method of the quadratic covariation $\langle X, Y \rangle_T$. In the next section this problem is tackled by establishing a combined approach of subsampling techniques to handle noise contamination and the synchronization method.

4 An estimation method in the presence of non-synchronicity and noise

In this chapter an estimator for the quadratic covariation of two Itô processes of the type introduced in Assumption 1 is established in the setting of non-synchronous discrete observations that are contaminated with microstructure noise.
In the first Section 4.1 a generalized multiscale estimator is presented whose construction is based on a convenient combination of the methods to deal with asynchronous sampling schemes from Chapter 3 and subsampling techniques to reduce the effect of observation noise that have been motivated in Section 2.2. The utter utility of the synchronization algorithm 3.1 will be perceived by a comparison of the resulting generalized multiscale estimator to another plausible combined method. We stress that our synchronization method and the approach to non-synchronous observations developed in Chapter 3 combined with subsampling techniques leads to a more efficient estimator than an extension of the "original" Hayashi-Yoshida estimator.
The asymptotic properties of our estimator are analyzed in Section 4.2. Balancing the errors due to noise and discretization, rate-optimality of the estimator is attained. We give a stable central limit theorem and the key ingredients for its proof are expounded. The detailed proof of stable weak convergence of the generalized multiscale estimator to a centred mixed Gaussian distribution with optimal rate is postponed to Section 4.3.

4.1 A generalized multiscale approach

The statistical model considered throughout this chapter is specified within the following assumptions:

Assumption 4.1 (**observed processes with noise**). *Given the deterministic observation schemes $\mathcal{T}^{X,n} = \{0 \leq t_0^{(n)} < t_1^{(n)} < \ldots < t_n^{(n)} \leq T\}$ and $\mathcal{T}^{Y,m} = \{0 \leq \tau_0^{(m)} < \tau_1^{(m)} < \ldots < \tau_m^{(m)} \leq T\}$ according to Assumption 2(b), two efficient processes X and Y that satisfy Assumption 1 are discretely observed at the times $\mathcal{T}^{X,n}$ and $\mathcal{T}^{Y,m}$ with additive observation noise:*

$$\tilde{X}_{t_i^{(n)}} = \int_0^{t_i^{(n)}} \mu_t^X \, dt + \int_0^{t_i^{(n)}} \sigma_t^X \, dB_t^X + \epsilon^X_{t_i^{(n)}} \;, 0 \leq i \leq n \;, \tag{4.1a}$$

$$\tilde{Y}_{\tau_j^{(m)}} = \int_0^{\tau_j^{(m)}} \mu_t^Y \, dt + \int_0^{\tau_j^{(m)}} \sigma_t^Y \, dB_t^Y + \epsilon^Y_{\tau_j^{(m)}} \;, 0 \leq j \leq m \;. \tag{4.1b}$$

Assumption 3 (microstructure noise). *The discrete microstructure noise processes*

$$\epsilon^X_{t_i^{(n)}}, \epsilon^Y_{\tau_j^{(m)}}, 0 \le i \le n, 0 \le j \le m \ .$$

are assumed to be centred i. i. d., independent of each other and independent of the efficient processes X and Y. We assume that the observation errors have finite fourth moments and denote the variances

$$\eta_X^2 = \mathbb{V}\mathrm{ar}\left(\epsilon^X_{t_1^{(n)}}\right) \ , \ \eta_Y^2 = \mathbb{V}\mathrm{ar}\left(\epsilon^Y_{\tau_1^{(m)}}\right) \ .$$

The noise variances are modeled not to depend on n and m.

We propose the following combined ex-post estimation method for the integrated covariance from noisy asynchronous observations. After applying Algorithm 3.1 to the observation times, the generalized multiscale estimator is defined as

$$\widehat{\langle X,Y\rangle}_T^{multi} = \sum_{i=1}^{M_N} \frac{\alpha^{opt}_{i,M_N}}{i} \sum_{j=i}^{N} \left(\tilde{X}_{g_j^{(N)}} - \tilde{X}_{l_{j-i+1}^{(N)}}\right)\left(\tilde{Y}_{\gamma_j^{(N)}} - \tilde{Y}_{\lambda_{j-i+1}^{(N)}}\right) \ . \tag{4.2}$$

This generalized multiscale estimator is a weighted sum of M_N one-scale subsampling estimators

$$\widehat{\langle X,Y\rangle}_T^{sub} = \frac{1}{i} \sum_{j=i}^{N} \left(\tilde{X}_{g_j^{(N)}} - \tilde{X}_{l_{j-i+1}^{(N)}}\right)\left(\tilde{Y}_{\gamma_j^{(N)}} - \tilde{Y}_{\lambda_{j-i+1}^{(N)}}\right) \tag{4.3}$$

with subsampling frequencies $i = 1, \ldots, M_N$ and weights α_{i,M_N} with $\sum \alpha_{i,M_N} = 1$. The optimal weights in the definition of the estimator (4.2) are determined later in (4.14). Owing to the aggregation of non-synchronous observations before applying sparse-sampling, subsampling and the multiscale approach, the resulting estimator has a conformable appearance as in the synchronous case (2.11a). Recall that in the synchronous setting $m = n$ and $t_j^{(n)} = \tau_j^{(m)}$ for all j, $g_j^{(N)} = \gamma_j^{(N)} = T_j^{(N)}$ and $l_{j-i+1}^{(N)} = \lambda_{j-i+1}^{(N)} = T_{j-i}^{(N)}$ holds, since $l_k^{(N)}$ and $\lambda_k^{(N)}$ denote the observation times preceding the minima of $\mathcal{H}^k$ and $\mathcal{G}^k$. Summation in the estimators starts with $j = i$ to avoid errors by the initial term and the convention $l_0^{(N)} = t_0^{(N)}$, $\lambda_0^{(N)} = \tau_0^{(N)}$.
Choosing $M_N = c_{multi} \cdot \sqrt{N}$ and $i_N = c_{sub} \cdot N^{2/3}$, both estimators above provide consistent and asymptotically unbiased estimators with convergence rate $N^{1/4}$ and $N^{1/6}$, respectively. We remark that the generalized multiscale estimator (4.2) differs from the other plausible Hayashi-Yoshida version of a multiscale estimator

$$\sum_{i=1}^{M_N} \frac{\beta^{opt}_{i,M_N}}{i} \sum_{j=i}^{n} \sum_{k \in \mathbb{Z}} \left(\tilde{X}_{t_j^{(n)}} - \tilde{X}_{t_{j-i}^{(n)}}\right)\left(\tilde{Y}_{\tau^{(m)}_{j+k\cdot i}} - \tilde{Y}_{\tau^{(m)}_{j+(k-1)\cdot i}}\right)\mathbb{1}_{\{\max(t_{j-i}^{(n)}, \tau^{(m)}_{j+(k-1)\cdot i}) < \min(t_j^{(n)}, \tau^{(m)}_{j+k\cdot i})\}} \tag{4.4}$$

and we state without proof that this estimator, which arises as natural Hayashi-Yoshida

multiscale estimator when, on the basis of (non-synchronized) observations of $\tilde{X}$ and $\tilde{Y}$, sparse-sample Hayashi-Yoshida estimators are averaged to one-scale subsample estimators and those extended to a weighted sum using different time lags, is consistent and asymptotically unbiased as well. Furthermore, it will attain the optimal rate of convergence. Nevertheless, we benefit from the data aggregation method and applying subsampling methods to already synchronized observations, since the variance of our estimator (4.2) is smaller than the one of this alternative estimator and we are able to find a feasible closed-form expression of the asymptotic variance.
The crucial difference between both approaches is that for the alternative method next- and previous-tick interpolation errors take place on sparse-sampling time intervals of order i/m whereas the interpolation errors of the generalized multiscale estimator (4.2) take place on intervals of order $1/N$ and "keep to the highest-frequency-scale". In particular the decomposition

$$\left(X_{g_j^{(N)}} - X_{l^{(N)}_{j-i_N+1}}\right) = \big(\underbrace{X_{g_j^{(N)}} - X_{T_j^{(N)}}}_{=\mathcal{O}_p(N^{-1/2})} + \underbrace{X_{T_j^{(N)}} - X_{T^{(N)}_{j-i}}}_{=\mathcal{O}_p((i/N)^{(1/2)})} + \underbrace{X_{T^{(N)}_{j-i}} - X_{l^{(N)}_{j-i_N+1}}}_{=\mathcal{O}_p(N^{-1/2})}\big)$$

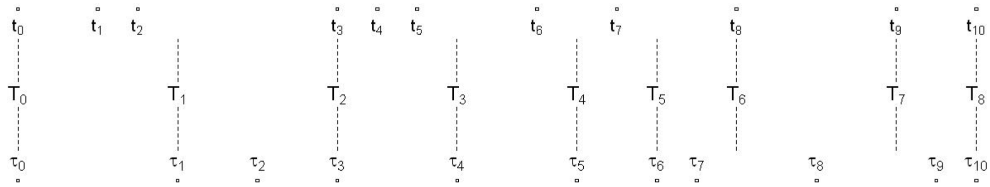
of the increments of X and analogously for Y, give a heuristic that the interpolation errors driving the so-called error due to non-synchronicity for the efficient process will asymptotically not affect the total discretization variance. The $T_j^{(N)}s, j = 0, \ldots, N$ denote the times defining the closest synchronous approximation that has been introduced in (3.4) and the stochastic orders are given for times instants of average order N^{-1}.

Example
For further motivation and to guarantee a better understanding of the combined estimation methods for noisy non-synchronous data, we take up again the example considered in Section 3.1.
The application of Algorithm 3.1 to the observation time schemes of the example has led to the sets $\mathcal{H}^i, \mathcal{G}^i, i = 0, \ldots, 8$ and $g_i, l_i, \gamma_i, \lambda_i, T_i$ listed in Figure 4.1. We illustrate the construction of a one-scale subsampling estimator with subsampling frequency $i = 3$. Denote X and Y the processes observed conforming to the sampling schemes of Figure 4.1. With the values given in the table we derive for the one-scale estimator with $i = 3$:

$$\begin{aligned}
&\frac{1}{3}\left(\sum_{j=3}^{8}\left(X_{g_j} - X_{l_{j-2}}\right)\left(Y_{\gamma_j} - Y_{\lambda_{j-2}}\right)\right)\\
&= \frac{1}{3}\left[\left(X_{t_6} - X_{t_0}\right)\left(Y_{\tau_4} - Y_{\tau_0}\right) + \left(X_{t_7} - X_{t_2}\right)\left(Y_{\tau_5} - Y_{\tau_1}\right) + \left(X_{t_8} - X_{t_3}\right)\left(Y_{\tau_6} - Y_{\tau_3}\right)\right.\\
&\quad \left. + \left(X_{t_8} - X_{t_5}\right)\left(Y_{\tau_8} - Y_{\tau_4}\right) + \left(X_{t_9} - X_{t_6}\right)\left(Y_{\tau_9} - Y_{\tau_5}\right) + \left(X_{t_{10}} - X_{t_7}\right)\left(Y_{\tau_{10}} - Y_{\tau_6}\right)\right]\\
&= \frac{1}{3}\Big[(\overrightarrow{X_{t_6} - X_{T_3}} + X_{T_3} - X_{T_0})(Y_{T_3} - Y_{T_0}) + (\overrightarrow{X_{t_7} - X_{T_4}} + X_{T_4} - X_{T_1} + \overleftarrow{X_{T_1} - X_{t_2}})\\
&\quad +(\overrightarrow{X_{t_8} - X_{T_5}} + X_{T_5} - X_{T_2})(Y_{T_5} - Y_{T_2}) + (X_{T_6} - X_{T_3}\\
&\quad +\overleftarrow{X_{T_3} - X_{t_5}})(\overrightarrow{Y_{\tau_8} - Y_{T_6}} + Y_{T_6} - Y_{T_3})
\end{aligned}$$

$$
\begin{aligned}
&+(X_{T_7}-X_{T_4}+\overleftarrow{X_{T_4}-X_{t_6}})(\overrightarrow{Y_{\tau_9}-Y_{T_7}}+Y_{T_7}-Y_{T_4})\\
&+(X_{T_8}-X_{T_5}+\overleftarrow{X_{T_5}-X_{t_7}})(Y_{T_8}-Y_{T_5})\Big]\\
=&\frac{1}{3}\Bigg[\Delta X_{T_1}\Delta Y_{T_1}+2\Delta X_{T_2}\Delta Y_{T_2}+3\sum_{k=3}^{6}\Delta X_{T_k}\Delta Y_{T_k}+2\Delta X_{T_7}\Delta Y_{T_7}+\Delta X_{T_8}\Delta Y_{T_8}\\
&+\sum_{k=1}^{8}\Delta X_{T_k}\sum_{r=(k-2)\vee 1}^{(k+2)\wedge 8}\left(1-\frac{r}{3}\right)\Delta Y_{T_r}+\sum_{k=1}^{8}\Delta Y_{T_k}\sum_{r=(k-2)\vee 1}^{(k+2)\wedge 8}\left(1-\frac{r}{3}\right)\Delta X_{T_r}\\
&+\overrightarrow{(X_{t_6}-X_{T_3})}(Y_{T_3}-Y_{T_0})+\overrightarrow{(X_{t_7}-X_{T_4})}(Y_{T_4}-Y_{T_1})+\overrightarrow{(X_{t_8}-X_{T_5})}(Y_{T_5}-Y_{T_2})\\
&+\overrightarrow{(Y_{\tau_8}-Y_{T_6})}(X_{T_6}-X_{T_3})+\overrightarrow{(Y_{\tau_9}-Y_{T_7})}(X_{T_7}-X_{T_4})\\
&+\overleftarrow{(X_{T_1}-X_{t_2})}(Y_{T_4}-Y_{T_1})+\overleftarrow{(X_{T_3}-X_{t_5})}(Y_{T_6}-Y_{T_3})+\overleftarrow{(X_{T_4}-X_{t_6})}(Y_{T_7}-Y_{T_4}\\
&+\overleftarrow{(X_{T_5}-X_{t_7})}(Y_{T_8}-Y_{T_5})\\
&+\overrightarrow{(Y_{\tau_8}-Y_{T_6})}\overleftarrow{(X_{T_3}-X_{t_5})}+\overrightarrow{(Y_{\tau_9}-Y_{T_7})}\overleftarrow{(X_{T_4}-X_{t_6})}\Bigg]\ .
\end{aligned}
$$

After inserting the conforming times according to the table in Figure 4.1, we have rewritten the increments separated in next-tick interpolations from times $T_i, i=0,\ldots,8$, previous-tick interpolated increments and increments over refresh time instants $T_i - T_{i-3}, i = 3,\ldots,8$. Previous-tick interpolations are marked with left-arrows and next-tick interpolations with right-arrows.
In the last equality we have multiplied out the increments and split the estimator into terms that only depend on the closest synchronous approximation, next-tick interpolations multiplied with increments over refresh time instants, previous-tick interpolations multiplied with increments over the synchronous approximation and next-tick times previous-tick interpolated increments. We denote $\Delta\cdot_i = \cdot_i - \cdot_{i-1}$ the backward difference operator as before.
Even if we suppose X and Y to have only Brownian components, the estimator for the example is biased anyway. This bias is yet only caused by boundary effects of the first i and last $(N-i)$ terms, that will be asymptotically negligible in the case that $i_N\to\infty, i_N/N\to 0$ as $N\to\infty$. The addends with cross terms of increments over refresh time intervals and all addends incorporating interpolated increments are centred in the zero-drift case. The sum of the cross terms is the leading term driving the asymptotic discretization variance of the synchronous part of the estimator. We will see that it is also leading for the overall discretization variance for the non-synchronous case. The reason is that there appear at most $3N$ terms with interpolation errors in the last illustration of the one-scale subsampling estimators and those interpolation errors are of order $\sqrt{\Delta T_i^{(N)}}$.
We calculate a one-scale estimator of the type (4.4) for the example visualized in Figure 4.1, too, to draw a comparison between both estimation approaches. The one-scale



i	0	1	2	3	4	5	6	7	8
$\mathcal{H}^i$	$\{t_0\}$	$\{t_1,t_2,t_3\}$	$\{t_3\}$	$\{t_4,t_5,t_6\}$	$\{t_6,t_7\}$	$\{t_7,t_8\}$	$\{t_8\}$	$\{t_9\}$	$\{t_{10}\}$
$\mathcal{G}^i$	$\{\tau_0\}$	$\{\tau_1\}$	$\{\tau_2,\tau_3\}$	$\{\tau_4\}$	$\{\tau_5\}$	$\{\tau_6\}$	$\{\tau_7,\tau_8\}$	$\{\tau_8,\tau_9\}$	$\{\tau_9,\tau_{10}\}$
l_i	t_0	t_0	t_2	t_3	t_5	t_6	t_7	t_8	t_9
g_i	t_0	t_3	t_3	t_6	t_7	t_8	t_8	t_9	t_{10}
λ_i	τ_0	τ_0	τ_1	τ_3	τ_4	τ_5	τ_6	τ_7	τ_8
γ_i	τ_0	τ_1	τ_3	τ_4	τ_5	τ_6	τ_8	τ_9	τ_{10}
T_i	$t_0{=}\tau_0$	τ_1	$t_3{=}\tau_3$	τ_4	τ_5	τ_6	t_8	t_9	$t_{10}{=}\tau_{10}$

Figure 4.1: Example for non-synchronous sampling design with the sets constructed by the synchronization algorithm, interpolated observation times occurring in the estimators and the synchronous approximation.

version of the estimator defined in (4.4) yields for $i=3$:

$$\frac{1}{3}\sum_{j=3}^{10}(X_{t_j}-X_{t_{j-3}})\sum_{k=0}^{3}(Y_{\tau_{3k+j}}-Y_{\tau_{3(k-1)+j}})\mathbb{1}_{\{\max(t_{j-3},\tau_{3(k-1)+j})<\min(t_j,\tau_{3k+j})\}}$$

$$\begin{aligned}
&= \frac{1}{3}\left[(X_{t_3}-X_{t_0})(Y_{\tau_3}-Y_{\tau_0})+(X_{t_4}-X_{t_1})(Y_{\tau_4}-Y_{\tau_1})\right.\\
&\quad +(X_{t_5}-X_{t_2})(Y_{\tau_5}-Y_{\tau_2})+(X_{t_6}-X_{t_3})(Y_{\tau_6}-Y_{\tau_3})\\
&\quad +(X_{t_7}-X_{t_4})(Y_{\tau_7}-Y_{\tau_4}+Y_{\tau_4}-Y_{\tau_1})+(X_{t_8}-X_{t_5})(Y_{\tau_8}-Y_{\tau_5}+Y_{\tau_5}-Y_{\tau_2})\\
&\quad \left.+(X_{t_9}-X_{t_6})(Y_{\tau_8}-Y_{\tau_5}+Y_{\tau_5}-Y_{\tau_2})+(X_{t_{10}}-X_{t_7})(Y_{\tau_{10}}-Y_{\tau_7}+Y_{\tau_7}-Y_{\tau_4})\right]\\
&= \frac{1}{3}\left[(X_{T_2}-X_{T_0})(Y_{T_2}-Y_{T_0})+(\overbrace{X_{t_4}-X_{T_2}}^{1}+X_{T_2}-X_{T_1}+X_{T_1}-X_{t_1})(Y_{T_3}-Y_{T_1})\right.\\
&\quad +(\overbrace{X_{t_5}-X_{T_2}}^{2}+X_{T_2}-X_{T_1}+X_{T_1}-X_{t_2})(Y_{T_4}-Y_{T_2}+Y_{T_2}-Y_{\tau_2})\\
&\quad +(\overbrace{X_{t_6}-X_{T_3}}^{3}+X_{T_3}-X_{T_2})(Y_{T_5}-Y_{T_2})\\
&\quad +(\overbrace{X_{t_7}-X_{T_4}}^{4}+X_{T_4}-X_{T_3}+\overbrace{X_{T_3}-X_{t_4}}^{1})(Y_{\tau_7}-Y_{T_5}+Y_{T_5}-Y_{T_1})\\
&\quad +(X_{T_6}-X_{T_3}+\overbrace{X_{T_3}-X_{t_5}}^{2})(Y_{\tau_8}-Y_{T_6}+Y_{T_6}-Y_{T_2}+Y_{T_2}Y_{\tau_2})\\
&\quad +(X_{T_7}-X_{T_4}+\overbrace{X_{T_4}-X_{t_6}}^{3})(Y_{\tau_9}-Y_{T_7}+Y_{T_7}-Y_{T_2})
\end{aligned}$$

$$
\begin{aligned}
&+ (X_{T_8} - X_{T_5} + \overbrace{X_{T_5} - X_{t_7}}^{4})(Y_{T_8} - Y_{T_3})] \\
= &\frac{1}{3}\left[(X_{T_2} - X_{T_0})(Y_{T_2} - Y_{T_0}) + (X_{T_3} - X_{T_1})(Y_{T_3} - Y_{T_1}) + (X_{T_4} - X_{T_2})(Y_{T_4} - Y_{T_2})\right. \\
&+(X_{T_5} - X_{T_2})(Y_{T_5} - Y_{T_2}) + (X_{T_5} - X_{T_3})(Y_{T_5} - Y_{T_3}) + (X_{T_6} - X_{T_4})(Y_{T_6} - Y_{T_4}) \\
&+(X_{T_7} - X_{T_5})(Y_{T_7} - Y_{T_5}) + (X_{T_8} - X_{T_5})(Y_{T_8} - Y_{T_5}) \\
&+(X_{T_1} - X_{t_2})(Y_{T_4} - Y_{T_2}) + (Y_{T_2} - Y_{\tau_2})(X_{T_4} - X_{t_2}) + (X_{T_5} - X_{t_7})(Y_{T_8} - Y_{T_5}) \\
&+(Y_{\tau_7} - Y_{T_5})(X_{t_7} - X_{T_3}) + (Y_{\tau_9} - Y_{T_7})(X_{T_7} - X_{T_5}) \\
&+(X_{T_5} - X_{T_4})(Y_{\tau_8} - Y_{T_5}) + (X_{T_6} - X_{T_5})(Y_{\tau_8} - Y_{T_6}) \\
&+(X_{T_1} - X_{t_1})(Y_{T_3} - Y_{T_1}) + (X_{T_3} - X_{t_4})(Y_{\tau_7} - Y_{T_3}) + (X_{T_4} - X_{t_5})(Y_{\tau_8} - Y_{T_4}) \\
&+(X_{T_5} - X_{t_6})(Y_{\tau_9} - Y_{T_5}) \\
&+(X_{T_6} - X_{T_4})(Y_{T_4} - Y_{\tau_2}) + (X_{t_7} - X_{T_3})(Y_{T_3} - Y_{T_1}) \\
&\left.+(X_{T_7} - X_{T_5})(Y_{T_5} - Y_{T_2}) + (X_{T_8} - X_{T_5})(Y_{T_5} - Y_{T_3})\right] \ .
\end{aligned}
$$

The increments indexed with same numerals are totalized and contribute addends that lead to the consistency of the estimator (4.4). The first term incorporating products of increments over joint sampling intervals equals

$$
\frac{1}{3}\left[\Delta X_{T_1}\Delta Y_{T_1} + 2\Delta X_{T_2}\Delta Y_{T_2} + 3\sum_{k=3}^{6}\Delta X_{T_k}\Delta Y_{T_k} + 2\Delta X_{T_7}\Delta Y_{T_7} + \Delta X_{T_8}\Delta Y_{T_8}\right]
$$

and addends with cross terms which again lead to a term that appears in the asymptotic variance as $i_N \to \infty$ and $N \to \infty$.
If drift terms are supposed to equal zero, the alternative estimator (4.4) for our example has the same expectation as the first estimator. The other addends occurring in the split sum differ from the ones we have obtained above for our one-scale subsampling estimator. We have separated all addends in the preceding equation in the ones that incorporate interpolated increments on the highest frequency as factors, increments over refresh time instants with lag 1, and products of increments that are both on a lower frequency. Those addends, asymptotically of order i_N/N, occurring when calculating an estimator of type (4.4), affect the variance also asymptotically. Thus, the variance of the estimator (4.4) is driven asymptotically by cross terms of refresh time increments as well as by interpolation errors on the subsampling time-scale.

4.2 Asymptotics of the estimators: Stable central limit theorems

In the following, we further illuminate the asymptotic characteristics as $N \to \infty, M_N \to \infty, M_N/N \to 0$ of the estimation methodology that has led to the generalized multiscale estimator (4.2) to cope with non-synchronous and noisy high-frequency observations when estimating the quadratic covariation of two Itô processes. We aim at establishing the seminal result of a feasible stable central limit theorem with optimal rate of convergence. To provide a feasible limit theorem, it is substantive to find a closed-form expression for the asymptotic variance on adequate assumptions on which the sequence of variances converges as $N \to \infty$. The issue of estimating the asymptotic variance will be suspended to Chapter 5.

A comprehensive analysis of the asymptotic estimation error necessitates a more elaborate screening of the conjunction of the data aggregation method and the joint sampling design $\left(\mathcal{T}^{X,n}, \mathcal{T}^{Y,m}\right)$. In particular, for a rigorous clarification of the asymptotic error due to noise and the cross term, both influenced by the i. i. d. observation errors at times $g_i^{(N)}, l_i^{(N)}, \gamma_i^{(N)}, \lambda_i^{(N)}$, we have to figure out the maxima $g_i^{(N)} = g_{i+1}^{(N)}$ of sets $\mathcal{H}^i$ that equal subsequent maxima and the maxima $g_i^{(N)} = l_{i+1}^{(N)}, g_i^{(N)} = l_{i+2}^{(N)}$ that are as well preceding observation times of minima of subsequent sets for $\mathcal{H}^i$ and analogously for $\mathcal{G}^i$. In the end we will learn that the numbers of repeating maxima $|\{g_i^{(N)} = g_{i+1}^{(N)}\}|$, $i = 1, \ldots, N-1$ and $|\{\gamma_i^{(N)} = \gamma_{i+1}^{(N)}\}|$ are vital for the asymptotic variance, the proof of a central limit theorem and to determine the influence of a certain degree of regularity of non-synchronous sampling schemes.

We start with an allocation of the maxima $g_i^{(N)}$ and $\gamma_i^{(N)}$, respectively, to four disjoint sets respective to one of the following four cases. We drop the upper indices of all observation times in the following. Denote $\gamma_{j,-}$ and $\gamma_{j,+}$ the observation times of $\tilde{Y}$ before and after γ_j. We illustrate the allocation of the observation times for $\mathcal{T}^{Y,m}$ and γ_j , $j = 1, \ldots, N-2$:

$$
\begin{aligned}
&① \; \gamma_j \le g_j && \Rightarrow \gamma_j \neq \gamma_{j+1} \; , \; \gamma_j = \lambda_{j+1} \; , \; \gamma_j \neq \lambda_{j+2} , \\
&② \; \gamma_j > g_j \; , \; \gamma_j \ge g_{j,+} && \Rightarrow \gamma_j = \gamma_{j+1} \; , \; \gamma_j \neq \lambda_{j+1} \; , \; \gamma_j = \lambda_{j+2} , \\
&③ \; \gamma_j > g_j \; , \; \gamma_j < g_{j,+} \; , \; \gamma_{j,+} > g_{j,+} && \Rightarrow \gamma_j \neq \gamma_{j+1} \; , \; \gamma_j \neq \lambda_{j+1} \; , \; \gamma_j = \lambda_{j+2} , \\
&④ \; \gamma_j > g_j \; , \; \gamma_j < g_{j,+} \; , \; \gamma_{j,+} \le g_{j,+} && \Rightarrow \gamma_j \neq \gamma_{j+1} \; , \; \gamma_j \neq \lambda_{j+1} \; , \; \gamma_j \neq \lambda_{j+2} , \; \gamma_{j,+} = \lambda_{j+2} .
\end{aligned}
$$

All observation times γ_i, λ_i are characterized through one of those four cases. Only sampling times distributed to case ② lead to repeated maxima $\gamma_i = \gamma_{i+1}$. In cases ①, ② and ③ a subsequent left-end point $\lambda_k, k = i+1$ or $k = i+2$ of observation time instants incorporated in the subsampling estimators is designated by γ_i. All λ_k, $k = 2, \ldots, N$, that are not maxima of previous sets $\mathcal{G}^{k-1}$ or $\mathcal{G}^{k-2}$ appear in an allocation of sampling times of the type ④, where $\lambda_{j+2} = \gamma_{j,+} \neq \gamma_l \, \forall l$. Recall that $\lambda_i \neq \lambda_k$ for all $i \neq k$ holds true.

If ② holds for γ_j with fixed $j \in \{1, \ldots, N-2\}$ and if

$$k := \arg \min_{k \in \{j, \ldots, N-1\}} (\gamma_k > g_k \, , \, \gamma_k \ge g_{k,+})$$

exists, then ② holds necessarily for one $g_l, l \in \{j+1, \ldots, k-1\}$ or $g_l = \gamma_l$. In Table 4.1 we list the cases and the relations for the sampling design of our prevenient example.

Definition 4.2.1 (degree of regularity of asynchronicity). *For $N \in \mathbb{N}$ and sets $\mathcal{H}^i, \mathcal{G}^i, i = 0, \ldots, N$ constructed from aggregated sampling schemes $\mathcal{T}^{X,n}, \mathcal{T}^{Y,m}$ that fulfill*

i	1	2	3	4	5	6	7	8
case X	②	①	③	③	②	①	①	①
case Y	①	①	①	①	①	③	④	①
relations X	$g_1 = g_2 = l_3$	$g_2 = l_3$	$g_3 = l_5$	$g_4 = l_6$	$g_5 = l_7$	$g_6 = l_7$	$g_7 = l_8$	–
relations Y	$\gamma_1 = \lambda_2$	$\gamma_2 = \lambda_3$	$\gamma_3 = \lambda_4$	$\gamma_4 = \lambda_5$	$\gamma_5 = \lambda_6$	$\gamma_6 = \lambda_8$	–	–

Table 4.1: Allocation of sampling times to cases ① – ④.

Assumption 4.1, define the following sequences of functions:

$$I_X^N(t) = \frac{T}{N} \sum_{g_j^{(N)} \le t} \mathbb{1}_{\{g_j^{(N)} = g_{j-1}^{(N)}\}} \ , \tag{4.5a}$$

$$I_Y^N(t) = \frac{T}{N} \sum_{\gamma_j^{(N)} \le t} \mathbb{1}_{\{\gamma_j^{(N)} = \gamma_{j-1}^{(N)}\}} \ , \tag{4.5b}$$

which describe the degree of regularity of asynchronicity between observation times $\mathcal{T}^{X,n}$ and $\mathcal{T}^{Y,m}$.

In the completely asynchronous case, we can directly conclude that $|I_X^N(t) - I_Y^N(t)| \le T/N$ for all $t \in [0,T]$ and one sequence suffices to reflect the regularity of the non-synchronous sampling schemes.

Assumption 4.2 (asymptotic degree of regularity of asynchronicity). *Assume that for the sequences of sampling schemes and for the sequences of functions I_X^N, I_Y^N defined in Definition 4.2.1, the following holds true:*

(i) $I_X^N(t) \to I_X(t), I_Y^N(t) \to I_Y(t)$ *as* $N \to \infty$*, where* $I_X(t), I_Y(t)$ *are continuously differentiable functions on* $[0,T]$.

(ii) For any null sequence (h_N), $h_N = \mathcal{O}(N^{-1})$

$$\frac{I_X^N(t+h_N) - I_X^N(t)}{h_N} \to I_X'(t) \ , \tag{4.6a}$$

$$\frac{I_Y^N(t+h_N) - I_Y^N(t)}{h_N} \to I_Y'(t) \tag{4.6b}$$

uniformly on $[0,T]$ *as* $N \to \infty$.

For synchronous settings and intermeshed sampling schemes that have been introduced in Section 3.2, the sequence of functions I_X^N, I_Y^N satisfy $I_X^N(t) \equiv 0 \equiv I_Y^N(t)$ for all N. The functions defined in Definition 4.2.1 are non-negative and bounded above by T. In Section 5.2 we will explore the degree of regularity of asynchronicity for independent

Poisson sampling schemes and explicitly deduce an asymptotic degree of regularity of asynchronicity. The term (asymptotic) degree of regularity of asynchronicity has been chosen since Assumption 4.2, that the sequences of difference quotients uniformly converge, are related to the boundedness of second derivatives if those exist. Assumption 4.2 will hold for all non-degenerate sequences where observation times conforming to one of the cases ① − ④ from above tend to be distributed according to some regular pattern. It is interesting and might seem surprising at first glance that the asymptotic variance of the estimator hinges on this asymptotic feature from Assumption 4.2 whereas, as indicated before, the asymptotic distribution or allocation of interpolation time instants will be asymptotically immaterial. The first circumstance is caused by the fact that for the construction of an estimator with Algorithm 3.1 as for the original Hayashi-Yoshida estimator (3.2), observed values of the processes at a certain observation time can appear twice. If there is observation noise, the number of observations allocated conforming to case ② has an impact on the asymptotics of the estimation method. The asymptotically vanishing influence of interpolation intervals for the combined method in the presence of noise in contrast to the non-noisy setting is due to the fact that interpolation steps take place on the time-scale of high-frequency observations, whereas lower-frequency sparse-sampled increments of the synchronous approximation are involved to reduce the distortion by noise contamination. This will be proved in detail below. We stress that for the preceding discussion we always suppose sampling schemes according to Assumption 2(b).
Essential for the asymptotic theory of the estimator (4.2) is the concept of the closest synchronous approximation. Recall the definition of the quadratic variation of time of the closest synchronous approximation from Definition 3.2.3. We impose as in Chapter 3, that the sequence of quadratic variations converges to a continuously differentiable function.

Assumption 4.3 (**asymptotic quadratic variation of time**). *Assume that for the sequences of sampling schemes and the times $T_i^{(N)}$ of the closest synchronous approximations and for the sequence of quadratic variations of time $G^N(t)$ defined in Definition 3.2.3, the following holds true:*

(i) $G^N(t) \to G(t)$ *as* $N \to \infty$, *where* $G(t)$ *is a continuously differentiable function on* $[0, T]$.

(ii) For any null sequence (h_N), $h_N = \mathcal{O}\left(N^{-1}\right)$

$$\frac{G^N(t + h_N) - G^N(t)}{h_N} \to G'(t) \tag{4.7}$$

uniformly on $[0, T]$ *as* $N \to \infty$..

(iii) The derivative $G'(t)$ *is bounded away from zero.*

Assumption 4.2 (i) and (ii) have already been part of Assumption 3.1 in Chapter 3, but Assumption 4.2 is weaker since it suffices to assume convergence of the quadratic

variation and we do not need to assume convergence of the (co-)variations including next- and previous-ticks to establish a stable central limit theorem. The auxiliary mild condition (iii) is additionally imposed to be able to define a time-change of the sampling design which is introduced below.
We have gathered all key ingredients and assumptions for a detailed analysis of the asymptotic estimation error and continue with the central result of this work in Theorem 4.1.

Theorem 4.1 (**Central limit theorem for the generalized multiscale estimator**). *On the Assumptions 1, 3, 4.1, 4.2 and 4.3, the generalized multiscale estimator* (4.2) *with noise-optimal weights* $\alpha_{i,M_N}^{opt} = (12i^2/M_N^3) - (6i/M_N^2)\,(1 + \mathcal{O}(1))$*, that are explicitly given in* (4.14)*, and* $M_N = c_{multi} \cdot \sqrt{N}$ *converges stably in law with optimal rate* $N^{1/4}$ *to a mixed Gaussian limiting distribution:*

$$N^{1/4}\left(\widehat{\langle X,Y\rangle}_T^{multi} - \langle X,Y\rangle_T\right) \overset{st}{\rightsquigarrow} \mathbf{N}\left(0, \mathbf{AVAR}_{multi}\right)$$

with the asymptotic variance

$$\begin{aligned}\mathbf{AVAR}_{multi} = c_{multi}^{-3}\, &\underbrace{\left(24 + 12\,\frac{I_X(T)+I_Y(T)}{T}\right)\eta_X^2\eta_Y^2}_{=\mathbf{AVAR}_n} + c_{multi}^{-1}\,\frac{12\eta_X^2\eta_Y^2}{5} \\ + c_{multi}\, &\underbrace{\frac{26}{35}T\int_0^T G'(t)(\sigma_t^X\sigma_t^Y)^2(1+\rho_t^2)\,dt}_{=\mathbf{AVAR}_{dis}^{multi}} \\ + c_{multi}^{-1}\, &\underbrace{\frac{12}{5}\left(\eta_Y^2\int_0^T(1+I_Y'(t))(\sigma_t^X)^2\,dt + \eta_X^2\int_0^T(1+I_X'(t))(\sigma_t^Y)^2\,dt\right)}_{=\mathbf{AVAR}_{cross}}\,.\end{aligned} \tag{4.8}$$

The weak convergence is proved to be stable with respect to the σ-algebra $\mathcal{F}$ associated with the efficient processes. As a side result, we also obtain a stable central limit theorem for a simpler one-scale subsampling estimator which is given for the sake of completeness in the next Corollary 4.2.2.

Corollary 4.2.2 (**Central limit theorem for the one-scale subsampling estimator**). *On the Assumptions 1, 3, 4.1 and 4.3, the one-scale subsampling estimator with subsampling frequency* $i_N = c_{sub} \cdot N^{2/3}$ *converges* $\mathcal{F}$*-stably in law with rate* $N^{1/6}$ *to a mixed Gaussian limiting distribution:*

$$N^{1/6}\left(\widehat{\langle X,Y\rangle}_T^{sub} - \langle X,Y\rangle_T\right) \overset{st}{\rightsquigarrow} \mathbf{N}\left(0, \mathbf{AVAR}_{sub}\right)\ , \tag{4.9}$$

with the asymptotic variance

$$\mathbf{AVAR}_{sub} = c_{sub}^{-2} \underbrace{4\eta_X^2\eta_Y^2}_{=\mathbf{AVAR}_{n,sub}} + c_{sub} \underbrace{\frac{2}{3}T\int_0^T G'(t)(\sigma_t^X\sigma_t^Y)^2(1+\rho_t^2)\,dt}_{=\mathbf{AVAR}_{dis,sub}} \,. \tag{4.10}$$

For the proof of Theorem 4.1, we split the total estimation error of the generalized multiscale estimator in three asymptotically uncorrelated parts:

$$\left(\widehat{\langle X,Y\rangle}_T^{multi} - \langle X,Y\rangle_T\right) = \underbrace{\sum_{i=1}^{M_N}\frac{\alpha_{i,M_N}^{opt}}{i}\sum_{j=i}^{N}\left(\epsilon_{g_j}^X - \epsilon_{l_{j-i+1}}^X\right)\left(\epsilon_{\gamma_j}^Y - \epsilon_{\lambda_{j-i+1}}^Y\right)}_{\text{error due to noise}}$$
$$+\underbrace{\sum_{i=1}^{M_N}\frac{\alpha_{i,M_N}^{opt}}{i}\sum_{j=i}^{N}\left(\left(X_{g_j}-X_{l_{j-i+1}}\right)\left(\epsilon_{\gamma_j}^Y - \epsilon_{\lambda_{j-i+1}}^Y\right) + \left(Y_{\gamma_j}-Y_{\lambda_{j-i+1}}\right)\left(\epsilon_{g_j}^X - \epsilon_{l_{j-i+1}}^X\right)\right)}_{\text{cross term}}$$
$$+\underbrace{\sum_{i=1}^{M_N}\frac{\alpha_{i,M_N}^{opt}}{i}\sum_{j=i}^{N}(X_{g_j}-X_{l_{j-i+1}})(Y_{\gamma_j}-Y_{\lambda_{j-i+1}}) - \langle X,Y\rangle_T}_{\text{discretization error}} \,.$$

For the one-scale subsampling estimator we follow the same ansatz. Since the observation errors are assumed to be centred i. i. d. and independent of the efficient processes X and Y, the three different error types in the above decomposition are uncorrelated if we ignore drift terms and else asymptotically uncorrelated.
The convergence rates and the orders of the errors have been derived in Bibinger [2011]. Here, we additionally focus on the asymptotic distribution of the estimation errors.
The error due to microstructure noise of the one-scale subsampling estimator has expectation zero and the variance yields

$$i_N^{-2}\sum_{j=i_N}^{N}\mathbb{E}\left[\left(\epsilon_{g_j}^X - \epsilon_{l_{j-i_N+1}}^X\right)^2\left(\epsilon_{\gamma_j}^Y - \epsilon_{\lambda_{j-i_N+1}}^Y\right)^2\right] = 4Ni_N^{-2}\eta_X^2\eta_Y^2 + \mathcal{O}\left(Ni_N^{-2}\right) \,.$$

We have used that the observation noise of $\tilde{X}$ and $\tilde{Y}$ are independent of each other by Assumption 3 and that $l_k \neq l_r$ for $k \neq r$, $\lambda_k \neq \lambda_r$ for $k \neq r$ and if $g_k = g_{k+1} \Rightarrow \gamma_k < \gamma_{k+1}\ \forall k$. The error due to noise is hence a sum of i. i. d. centred random variables and the standard central limit theorem applies.
The analysis for the generalized multiscale estimator becomes more involved. In Subsection 4.3.1, we further decompose the error due to noise in a main part of order $N^{1/2}M_N^{-3/2}$ and two terms due to end-effects of orders $M_N^{-1/2}$, where all three terms are uncorrelated. Propositions 4.3.1 and 4.3.2 give limit theorems for these terms. Asymptotic normality

for the error due to noise holds conditionally and unconditionally on the paths of the efficient processes.
The asymptotic variance of the one-scale estimator is simpler compared to the generalized multiscale estimator for two reasons. The error due to noise does not depend on any further influence of the sampling schemes except the number of constructed sets N and the cross terms are, in contrast to the multiscale case, asymptotically negligible since

$$\mathbb{E}\left[\left(i_N^{-1}\sum_{j=i_N}^{N}\left(X_{g_j}-X_{l_{j-i_N+1}}\right)\left(\epsilon^Y_{\gamma_j}-\epsilon^Y_{\lambda_{j-i_N+1}}\right)+\left(Y_{\gamma_j}-Y_{\lambda_{j-i_N+1}}\right)\left(\epsilon^X_{g_j}-\epsilon^X_{l_{j-i_N+1}}\right)\right)^2\right]$$
$$=i_N^{-2}\sum_{j=i_N}^{N}\left(\int_{l_{j-i_N+1}}^{g_j}(\sigma_t^X)^2\,dt\;2\eta_Y^2+\int_{\lambda_{j-i_N+1}}^{\gamma_j}(\sigma_t^Y)^2\,dt\;2\eta_X^2\right)+\mathcal{o}\left(i_N^{-2}\right)$$
$$=\mathcal{O}\left(i_N^{-2}\right)=\mathcal{o}(1)\,.$$

For the generalized multiscale estimator instead the cross term is of order $M_N^{-1/2}$ and will have effect upon the asymptotic distribution. In Proposition 4.3.15 a limit theorem is stated where the weak convergence also holds conditionally and unconditionally on the paths of the efficient processes. The asymptotic variance $\mathbf{AVAR}_{cross}$ includes the influence of the asymptotic degree of regularity of asynchronicity.
The discretization error will engage most of the theory in Section 4.3. Nevertheless, we benefit from the fact that for our combined methods and the closest synchronous approximation the error due to non-synchronicity is of smaller asymptotic order. The discretization error of the generalized multiscale estimator is of order $M_N^{1/2}N^{-1/2}$ and that of the one-scale estimator of order $i_N^{1/2}N^{-1/2}$. We prepare the proof of stable central limit theorems for those, that are given in Propositions 4.3.4 and 4.3.3, in the following.
For the moment, we observe that there is a trade-off between the error due to noise and the discretization error for both estimators. For the generalized multiscale estimator these are of orders $N^{1/2}M_N^{-3/2}$ and $M_N^{1/2}N^{-1/2}$, respectively. Remaining other terms are of orders $M_N^{-1/2}$. Thus, choosing $M_N=c_{multi}\cdot N^{1/2}$, the total estimation error is minimized and of order $M_N^{-1/2}=N^{-1/4}$ which constitutes the optimal rate of convergence in Theorem 4.1. The weak convergence of the discretization error is proved to be stable, so it converges jointly in law with every bounded $\mathcal{F}$-measurable random variable defined on the same probability space. Since the asymptotic normality of the cross term and the error due to noise holds both, conditionally and unconditionally given the efficient processes, and the discretization error is independent of ϵ^X and ϵ^Y we can apply Theorem 1.9 to the sum, namely the overall estimation error. It is adapted with respect to $\mathcal{A}_j=\sigma\left(\epsilon^X_{t_k},t_k<T_{j+1},\epsilon^Y_{\tau_k},\tau_k<T_{j+1},\mathcal{F}_{T_j}\right)$ where $\mathcal{F}$ is associated with the efficient processes. The asymptotic variance is the sum of those of the single addends since all addends are uncorrelated. The conditional convergence given the paths of the efficient process is stronger than $\mathcal{F}$-stable convergence. Therefore, the stable weak convergence result in Theorem 4.1 is deduced. With the Cramér-Wold device 1.5 joint normality and

asymptotic independence of the different errors can be concluded.
This is likewise for the one-scale estimator and Corollary 4.2.2. Choosing the subsampling frequency $i_N = c_{sub} \cdot N^{2/3}$ balances the variance of the error due to noise which is of order Ni^{-2} and the discretization variance of order iN^{-1} to be of order $N^{-1/3}$. In conclusion, the overall mean square error is of order $N^{-1/3}$ and with rate $N^{1/6}$ the one-scale subsampling estimator is asymptotically Gaussian distributed with the asymptotic variance $\mathbf{AVAR}_{sub}$, but cannot attain the optimal convergence rate.
Next, we set up the error due to discretization of a one-scale subsampling estimator with subsampling frequency $i \in \mathbb{N}$ for the detailed analysis of the asymptotic discretization errors of the one-scale (4.3) and the generalized multiscale estimator (4.2).

$$
\begin{aligned}
&\frac{1}{i}\sum_{j=i}^{N}\left(X_{g_j} - X_{l_{j-i+1}}\right)\left(Y_{\gamma_j} - Y_{\lambda_{j-i+1}}\right) - \langle X, Y\rangle_T \\
&= \frac{1}{i}\sum_{j=i}^{N}\left(X_{T_j} - X_{T_{j-i}}\right)\left(Y_{T_j} - Y_{T_{j-i}}\right) - \langle X, Y\rangle_T \\
&+ \frac{1}{i}\sum_{j=i}^{N}\Big[\left(X_{g_j} - X_{T_j}\right)\left(Y_{T_j} - Y_{T_{j-i}}\right) + \left(X_{g_j} - X_{T_j}\right)\left(Y_{T_{j-i}} - Y_{\lambda_{j-i+1}}\right) \\
&\quad + \left(X_{T_{j-i}} - X_{l_{j-i+1}}\right)\left(Y_{\gamma_j} - Y_{T_j}\right) + \left(X_{T_{j-i}} - X_{l_{j-i+1}}\right)\left(Y_{T_j} - Y_{T_{j-i}}\right) \\
&\quad + \left(Y_{\gamma_j} - Y_{T_j}\right)\left(X_{T_j} - X_{T_{j-i}}\right) + \left(Y_{T_{j-i}} - Y_{\lambda_{j-i+1}}\right)\left(X_{T_j} - X_{T_{j-i}}\right)\Big] \\
&= \frac{1}{i}\sum_{j=i}^{N}\left(\sum_{k=j-i+1}^{j}\Delta X_{T_k}\right)\left(\sum_{k=j-i+1}^{j}\Delta Y_{T_k}\right) - \langle X, Y\rangle_T \\
&+ \frac{1}{i}\sum_{j=i}^{N-1}\left[\left(X_{g_j} - X_{T_j}\right)\left(\sum_{k=j-i+1}^{j}\Delta Y_{T_k}\right) + \left(X_{g_j} - X_{T_j}\right)\left(Y_{T_{j-i}} - Y_{\lambda_{j-i+1}}\right)\right. \\
&\quad + \left(X_{T_{j-i}} - X_{l_{j-i+1}}\right)\left(Y_{\gamma_j} - Y_{T_j}\right) + \left(X_{T_j} - X_{l_{j+1}}\right)\left(\sum_{k=j-i+1}^{j}\Delta Y_{T_k}\right) \\
&\quad \left. + \left(Y_{\gamma_j} - Y_{T_j}\right)\left(\sum_{k=j-i+1}^{j}\Delta X_{T_k}\right) + \left(Y_{T_j} - Y_{\lambda_{j+1}}\right)\left(\sum_{k=j-i+1}^{j}\Delta X_{T_k}\right)\right] + \mathcal{O}_p\left(N^{-1}\right) \\
&= \frac{1}{i}\sum_{j=i}^{N}\left(\sum_{k=j-i+1}^{j}\Delta X_{T_k}\Delta Y_{T_k} + \sum_{\substack{l \neq r \\ l,r \in \{j-i+1,\ldots,j\}}}\Delta X_{T_l}\Delta Y_{T_r}\right) - \langle X, Y\rangle_T \\
&+ \frac{1}{i}\sum_{j=i}^{N-1}\left[\left(X_{g_j} - X_{T_j}\right)\left(\sum_{k=j-i+1}^{j}\Delta Y_{T_k}\right) + \left(Y_{\gamma_j} - Y_{T_j}\right)\left(\sum_{k=j-i+1}^{j}\Delta X_{T_k}\right)\right. \\
&\quad \left. + \Delta X_{T_{j+1}}\left(\sum_{k=j-i+1}^{j}\left(Y_{T_k} - Y_{\lambda_{k+1}}\right)\right) + \Delta Y_{T_{j+1}}\left(\sum_{k=j-i+1}^{j}\left(X_{T_k} - X_{l_{k+1}}\right)\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
&+ \mathcal{O}_p\left(i^{-1}N^{-\frac{1}{2}}\right) + \mathcal{O}_p\left(N^{-1}\right) \\
=& \sum_{j=1}^{N}\left(\Delta X_{T_j}\sum_{l=1}^{i\wedge j}\left(1-\frac{l}{i}\right)\Delta Y_{T_{j-l}} + \Delta Y_{T_j}\sum_{l=1}^{i\wedge j}\left(1-\frac{l}{i}\right)\Delta X_{T_{j-l}}\right) \\
&+ \mathcal{O}_p\left(iN^{-1}\right) + \mathcal{O}_p\left(N^{-1/2}\right) \\
&+ \mathcal{O}_p\left(i^{-1/2}N^{-1/2}\right) + \mathcal{O}_p\left(N^{-1}\right) + \mathcal{O}_p\left(i^{-1}N^{-1/2}\right)
\end{aligned}
$$

In this calculation we have written the total discretization error of a one-scale estimator as the sum of a discretization error of the closest synchronous approximation

$$
\sum_{j=1}^{N}\left(\Delta X_{T_j}\sum_{l=1}^{i\wedge j}\left(1-\frac{l}{i}\right)\Delta Y_{T_{j-l}} + \Delta Y_{T_j}\sum_{l=1}^{i\wedge j}\left(1-\frac{l}{i}\right)\Delta X_{T_{j-l}}\right) + \mathcal{O}_p\left(iN^{-1}\right) + \mathcal{O}_p\left(N^{-1/2}\right) \;, \tag{4.11}
$$

and an error due to the lack of synchronicity leading to the terms whose orders in probability are given in the last line. The subsampling estimator of the synchronous part has been split into

$$
\sum_{i=1}^{N}\Delta X_{T_i}\Delta Y_{T_i} - \langle X,Y\rangle_T = \mathcal{O}_p\left(N^{-1/2}\right)
$$

and the leading term of order $i^{1/2}N^{-1/2}$ that will drive the asymptotic distribution and is considered in the proof of a central limit theorem in the sequel. For all addends $i \le j \le N-i$, there are $(i-l)$ addends $\Delta X_{T_j}\Delta Y_{T_{j-l}}$ and $\Delta Y_{T_j}\Delta X_{T_{j-l}}$ appearing in the inner sums leading to the above given term. The remainder term of order iN^{-1} emerges from the $2i$ boundary addends.
In contrast to the synchronized realized covariance estimator considered in Chapter 3 in the non-noisy setting, the error due to non-synchronicity is asymptotically negligible here what we have heuristically motivated above. This is not true for the alternative estimator mentioned above and our decomposition of the total discretization error is only adequate here since for our method $\sqrt{N/i_N}$ times the error due to asynchronicity converges to zero in probability. Unlike the addends in the decomposition of the synchronized realized covariance estimator in Chapter 3, the error D_T^N of the closest synchronous approximation and the remaining error A_T^N for the Brownian parts of X and Y are not uncorrelated any more for the subsampling estimators. In particular,

$$
\begin{aligned}
\mathbb{E}\left[\left(D_T^N + A_T^N\right)^2\right] &= \mathbb{E}\left(D_T^N\right)^2 + 2\mathbb{E}\left[D_T^N A_T^N\right] + \mathbb{E}\left(A_T^N\right)^2 \\
&= \mathcal{O}\left(iN^{-1}\right) + \mathcal{O}\left(N^{-1}\right) + \mathcal{O}\left(i^{-1}N^{-1}\right)
\end{aligned}
$$

holds.
Since next-tick and previous-tick interpolations are only correlated at the same refresh

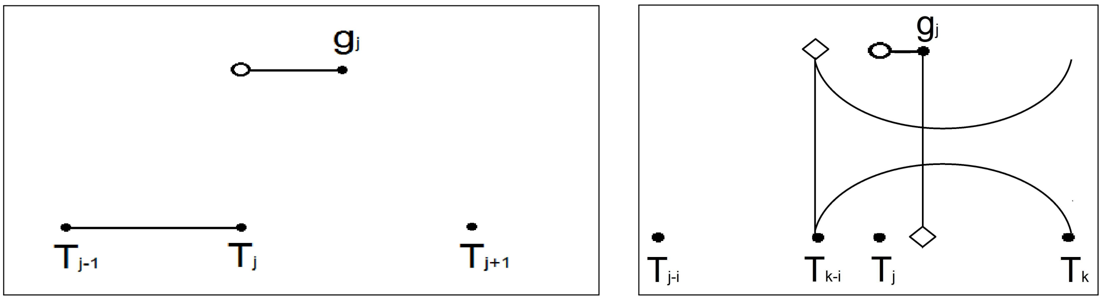


Figure 4.2: In the left part it is visualized that next-tick interpolation errors are uncorrelated to products of synchronous increments at the highest sampling frequency. This is not the case for lower-frequency sampling in the right part.

time $T_i, i = 1, \ldots, N-1$, we have applied an index shift in the sum above. End-effects lead to a term of order N^{-1} in probability and the addends incorporating next-tick multiplied with previous-tick interpolations are of order $i^{-1}N^{-1/2}$. The Brownian increments of those are uncorrelated and also uncorrelated to D_T^N. Hence, we have obtained

$$
\begin{aligned}
A_T^N = \frac{1}{i}\sum_{j=i}^{N-1} \Bigg[&\left(X_{g_j} - X_{T_j}\right)\left(\sum_{k=j-i+1}^{j} \Delta Y_{T_k}\right) + \left(Y_{\gamma_j} - Y_{T_j}\right)\left(\sum_{k=j-i+1}^{j} \Delta X_{T_k}\right) \\
&+ \Delta X_{T_{j+1}}\left(\sum_{k=j-i+1}^{j}\left(Y_{T_k} - Y_{\lambda_{k+1}}\right)\right) + \Delta Y_{T_{j+1}}\left(\sum_{k=j-i+1}^{j}\left(X_{T_k} - X_{l_{k+1}}\right)\right)\Bigg] \\
&+ \mathcal{O}_p\left(N^{-1}\right) + \mathcal{O}_p\left(i^{-1}N^{-1/2}\right) . \qquad (4.12)
\end{aligned}
$$

The leading term of A_T^N is rigorously treated in Section 4.3 and will be proved to be of order $i^{-1/2}N^{-1/2}$. At this point, we illustrate why D_T^N and A_T^N are in opposite to the estimator for non-noisy observations not uncorrelated for the subsampling case and prove that $\mathbb{E}\left[D_T^N A_T^N\right] = \mathcal{O}\left(N^{-1}\right) = o\left(iN^{-1}\right)$. Assume that X and Y are efficient processes according to Assumption 1 but without drift terms. In the left part of Figure 4.2 it can be seen that $\mathbb{E}\left[(X_{g_j} - X_{T_j})(\Delta Y_{T_j})^2 \Delta X_{T_j}\right] = 0$ as well as $\mathbb{E}\left[(X_{g_j} - X_{T_j})\Delta Y_{T_j}\Delta X_{T_{j+1}}\Delta Y_{T_{j+1}}\right] = 0$ holds. This has been the reason for the error of the closest synchronous approximation and the remaining discretization error due to non-synchronicity to be uncorrelated for the synchronized realized covariance estimator in Chapter 3. However, the right part of Figure 4.2 illustrates that $\mathbb{E}\left[D_T^N A_T^N\right] \neq 0$ in general for the subsampling case. In particular, $(X_{g_j} - X_{T_{j-i}})(Y_{T_j} - Y_{T_{j-i}})$ and $(X_{T_l} - X_{T_{l-i}})(Y_{T_l} - Y_{T_{l-i}})$ are uncorrelated if $|l-j| > i$. Otherwise, if without loss of generality $l \geq j$,

$$
\begin{aligned}
&\mathbb{E}\left[(X_{g_j} - X_{T_{j-i}})(Y_{T_j} - Y_{T_{j-i}})(X_{T_l} - X_{T_{l-i}})(Y_{T_l} - Y_{T_{l-i}})\right] \\
&= \mathbb{E}\left[(X_{g_j} - X_{T_{j-i}})(Y_{T_j} - Y_{T_{j-i}})(X_{T_j} - X_{T_{l-i}})(Y_{T_l} - Y_{T_j})\right]
\end{aligned}
$$

$$
\begin{aligned}
&+\mathbb{E}\left[(X_{g_j}-X_{T_{j-i}})(Y_{T_j}-Y_{T_{j-i}})(X_{T_l}-X_{T_j})(Y_{T_j}-Y_{T_{l-i}})\right]\\
&=\mathbb{E}\left[\int_{T_j}^{g_j}\rho_t\sigma_t^X\sigma_t^Y\,dt\right]\mathbb{E}\left[\int_{T_{l-i}}^{T_j}\rho_t\sigma_t^X\sigma_t^Y\,dt\right]+\mathbb{E}\left[\int_{T_j}^{g_j}(\sigma_t^X)^2\,dt\right]\mathbb{E}\left[\int_{T_{l-i}}^{T_j}(\sigma_t^Y)^2\,dt\right]\\
&=\mathcal{O}\left(iN^{-2}\right)\ .
\end{aligned}
$$

Hence, $\mathbb{E}\left[D_T^N A_T^N\right]$ is i^{-2} times the sum of expectations of next-tick interpolated increments, each multiplied with i addends of D_T^N, and analogous terms for all previous-tick interpolated increments. The total sum is then of order $i^{-2}\,iN\,iN^{-2}=N^{-1}$.

Proposition 4.2.3. *In the proof of a central limit theorem for the discretization error of the closest synchronous approximation $T_k^{(N)}$, $k=0,\dots,N$, of our generalized multiscale estimator* (4.2) *on the Assumptions 4.1 and 4.2, we can additionally, without loss of further generality, assume that*

$$
\sum_{k=1}^{N}\left(\Delta T_k^{(N)}-\frac{T}{N}\right)^2=o\left(N^{-1}\right)\ . \tag{4.13}
$$

Remark 4.1. *From Assumptions 1 and 4.1, we can deduce directly that the sum above is at most of order N^{-1}. The stronger assertion, that the closest synchronous approximation defined by the times $T_k^{(N)}$, $k=0,\dots,N$ introduced in* (3.4) *is close to equidistant sampling in the sense that the sum above is of smaller asymptotic order than N^{-1}, is derived by the concept of a time-change in the asymptotic quadratic variation of time from Assumptions 3.1 and 4.3. This concept has been presented in Zhang [2006] for the univariate multiscale approach and carries over to the synchronous multivariate case.*
On the Assumption 4.3, a transformation g can be defined that maps the refresh times $T_k^{(N)}$ to values $g\left(T_k^{(N)}\right)$, so that (4.13) *holds true for the transformed synchronous observation scheme. Thanks to the fact that the corresponding time-changed processes $L_{g(t)}$ and $M_{g(t)}$ fulfill Assumption 1 again and the transformed observation scheme Assumption 4.1, we are able to prove a central limit theorem for the time-changed version of the discretization error if* (4.13) *does not hold.*
Since the resulting asymptotic variance will be invariant under the transformation g, the central limit theorem will analogously hold true for the original sampling scheme. Hence, no further restriction has to be made when assuming (4.13). *Conditions* (1.3a) *and* (1.3b) *in Theorem 1.6 to prove stability of the weak convergence can be shown under the original underlying synchronous sampling scheme induced by the $T_k^{(N)}$, $k=0,\dots,N$, since the convergence of the covariations to zero will be proved on the Assumptions 1 and 4.1 and, hence,* (1.3a) *and* (1.3b) *for the transformed versions are equivalent to those for the original unchanged time-scale.*

Proof. From Assumption 4.1, the asymptotic quadratic variation of time $G(t)$ is Lipschitz continuous if it exists. Suppose Assumption 4.3 holds true. Define the mapping g :

$[0,T] \to [0,T^*]$ by

$$g(t) := \frac{T^*}{T} \int_0^t \left(G'(s)\right)^{-1} ds$$

with $T^* = G(T)$. This is an increasing Lipschitz continuous function for which

$$\tilde{G}(g(t)) := \lim_{N\to\infty} \frac{N}{T^*} \sum_{g(T_k^{(N)}) \le g(t)} \left(g(T_k^{(N)}) - g(T_{k-1}^{(N)})\right)^2 \ , \, t \in [0,T]$$

exists and $\tilde{G}'(g(t)) = (T/T^*)G'(t)g'(t)$ almost everywhere on $[0,T]$ (cf. Lemma 1 in Zhang [2006]).
Since $\sum_{k=1}^N \Delta T_k^{(N)} = T + \mathcal{O}(N^{-1})$, the left-hand side of (4.13) multiplied with N/T can be written

$$\frac{N}{T} \sum_{k=1}^N \left(\Delta T_k^{(N)} - \frac{T}{N}\right)^2 = \frac{N}{T} \sum_{k=1}^N \left(\Delta T_k^{(N)}\right)^2 - T + \mathcal{O}(N^{-1}) \ .$$

The expression on the right-hand side converges to zero if $G(T) = T$ and then (4.13) is implied.
For the above choice of g, it holds true that $\tilde{G}(T^*) = T^*$ for the time transformed asymptotic quadratic variation of time. $L_{g(t)}$ and $M_{g(t)}$ satisfy Assumption 1 again and the transformed synchronous sampling scheme $g(T_k^{(N)})$, $k = 0, \ldots, N$ Assumptions 4.1 and 4.3 with (4.13).
The asymptotic variance in (4.20) below is invariant under the transformation g. $\square$

The Cramér-Wold device 1.5 provides a connection between the weak convergence of a random vector and all its one-dimensional projections. We use this connection to start with a limit theorem for a finite-dimensional vector of discretization errors of the type (4.11) and derive the weak convergence of the weighted sum from the Cramér-Wold device. In the last step it is shown that the limit theorem can be extended to the case of infinite sums and, hence, stable convergence of the discretization error of the generalized multiscale estimator is ensured.

4.3 Proof of the stable central limit theorem

In the following, we mostly omit the superscripts (N) of the time points $g_i^{(N)}, \gamma_i^{(N)}, l_i^{(N)}$, $\lambda_i^{(N)}$, $i \in \{0, \ldots, N\}$.

4.3.1 Error due to noise and choosing the weights

The error due to microstructure noise of the generalized multiscale estimator is given by

$$\sum_{i=1}^{M_N}\frac{\alpha_{i,M_N}}{i}\sum_{j=i}^{N}\left(\epsilon^X_{g_j}-\epsilon^X_{l_{j-i+1}}\right)\left(\epsilon^Y_{\gamma_j}-\epsilon^Y_{\lambda_{j-i+1}}\right)=\sum_{i=1}^{M_N}\frac{\alpha_{i,M_N}}{i}\Big(\sum_{j=1}^{N}\left(\epsilon^X_{g_j}\epsilon^Y_{\gamma_j}+\epsilon^X_{l_j}\epsilon^Y_{\lambda_j}\right)$$
$$-\sum_{j=i}^{N}\left(\epsilon^X_{g_j}\epsilon^Y_{\lambda_{j-i+1}}+\epsilon^Y_{\gamma_j}\epsilon^X_{l_{j-i+1}}\right)-\sum_{j=1}^{i-1}\epsilon^X_{g_j}\epsilon^Y_{\gamma_j}-\sum_{j=N-i+1}^{N}\epsilon^X_{l_j}\epsilon^Y_{\lambda_j}\Big).$$

Additionally to the standardization condition

$$\sum_{i=1}^{M_N}\alpha_{i,M_N}=1\ , \tag{C1}$$

that is necessary for asymptotic unbiasedness and consistency, we now impose the auxiliary condition

$$\sum_{i=1}^{M_N}\frac{\alpha_{i,M_N}}{i}=0\ , \tag{C2}$$

on the weights which assures that the leading term of the noise error equals zero and hence there remain three uncorrelated addends in the error induced by microstructure noise.

Proposition 4.3.1. *Let Assumptions 4.1 and 4.2 on the observation times and Assumption 3 on the observation errors hold true. The asymptotic variance of the term*

$$-\sum_{i=1}^{M_N}\frac{\alpha_{i,M_N}}{i}\sum_{j=i}^{N}\left(\epsilon^X_{g_j}\epsilon^Y_{\lambda_{j-i+1}}+\epsilon^Y_{\gamma_j}\epsilon^X_{l_{j-i+1}}\right)$$

is minimized by the weights

$$\alpha^{opt}_{i,M_N}=\left(\frac{12i^2}{(M_N^3-M_N)}-\frac{6i}{(M_N^2-1)}-\frac{6i}{(M_N^3-M_N)}\right)=\frac{12i^2}{M_N^3}-\frac{6i}{M_N^2}\,(1+\mathcal{O}(1)) \tag{4.14}$$

as $M_N,N\to\infty$ and $M_N/N\to 0$ with $N=\mathcal{O}\left(M_N^4\right)$. The following asymptotic normality result holds true:

$$\sqrt{\frac{M_N^3}{N}}\left(\sum_{i=1}^{M_N}\frac{\alpha^{opt}_{i,M_N}}{i}\sum_{j=i}^{N}\left(\epsilon^X_{g_j}\epsilon^Y_{\lambda_{j-i+1}}-\epsilon^Y_{\gamma_j}\epsilon^X_{l_{j-i+1}}\right)\right)\rightsquigarrow\mathbf{N}\left(0\,,\,\mathbf{AVAR}_n\right)\ , \tag{4.15}$$

where the asymptotic variance

$$\mathbf{AVAR}_n=(24+12\kappa_n)\eta_X^2\eta_Y^2 \tag{4.16}$$

includes a constant $0 \leq \kappa_n \leq 1$ *depending on the asymptotic degree of regularity of asynchronicity which has been introduced in Definition 4.2.1 and Assumption 4.2. In the synchronous case* $\kappa_n = 0$ *holds. In particular, the constant* κ_n *equals*

$$\kappa_n = \frac{I_X(T) + I_Y(T)}{T} \tag{4.17}$$

where the functions I_X *and* I_Y *are defined in Assumption 4.2. The weak convergence also holds true conditionally given the paths of the efficient processes.*

Proof. Since the observation errors emanating from microstructure frictions are assumed to be centred i. i. d. and independent for both processes, the term is centred and we illustrate it in the way

$$-\sum_{i=1}^{M_N} \frac{\alpha_{i,M_N}}{i} \sum_{j=i}^{N} \left(\epsilon^X_{g_j}\epsilon^Y_{\lambda_{j-i+1}} + \epsilon^Y_{\gamma_j}\epsilon^X_{l_{j-i+1}}\right) = -\sum_{j=1}^{N} \sum_{i=1}^{M_N \wedge j} \frac{\alpha_{i,M_N}}{i} \Big(\epsilon^X_{g_j}\epsilon^Y_{\lambda_{j-i+1}}$$
$$(\mathbb{1}_{\{g_j = g_{j+1}\}} + \mathbb{1}_{\{g_j \neq g_{j+1}\}}) + \epsilon^Y_{\gamma_j}\epsilon^X_{l_{j-i+1}}(\mathbb{1}_{\{\gamma_j = \gamma_{j+1}\}} + \mathbb{1}_{\{\gamma_j \neq \gamma_{j+1}\}})\Big) .$$

For fixed i the addends of the inner sum are uncorrelated because all minima of sets $\mathcal{H}^i$ and $\mathcal{G}^i$, respectively, constructed by Algorithm 3.1 are different observation times. Consecutive maxima can be the same observation times instead, so that the inner sums are 2-dependent random variables. Thus, the variance is given by

$$\mathbb{V}\mathrm{ar}\left(\sum_{j=1}^{N} \frac{\alpha_{i,M_N}}{i} \sum_{i=1}^{M_N \wedge j} \left(\epsilon^X_{g_j}\epsilon^Y_{\lambda_{j-i+1}} - \epsilon^Y_{\gamma_j}\epsilon^X_{l_{j-i+1}}\right)\right)$$
$$= \sum_{j=1}^{N} \sum_{i=1}^{M_N \wedge j} \left(\frac{\alpha_{i,M_N}}{i}\right)^2 2\eta_X^2\eta_Y^2 + \sum_{j=1}^{N} \sum_{i=1}^{M_N^*(j)} \frac{\alpha_{i,M_N}\alpha_{i+1,M_N}}{i(i+1)} \eta_X^2\eta_Y^2 (1 - \mathbb{1}_{\{g_j \neq g_{j+1}\,,\,\gamma_j \neq \gamma_{j+1}\}}) ,$$

where $M^*(j) = (M_N - 1) \wedge (j-1)$. We can also rewrite the last term using the equality

$$(1 - \mathbb{1}_{\{g_j \neq g_{j+1}\,,\,\gamma_j \neq \gamma_{j+1}\}}) = (1 - \mathbb{1}_{\{\max(g_i,\gamma_i) < \min(g_{i,+},\gamma_{i,+})\}}) = \mathbb{1}_{\{\max(g_i,\gamma_i) \geq \min(g_{i,+},\gamma_{i,+})\}} .$$

This relation can be tracked with the considerations on observation schemes in Section 4.2. In fact, the autocorrelations to lag one of the addends in the second term are zero unless we are in case ② for the observation times of X or Y (cf. the discussion in Section 4.2).

In the following, we choose specific weights that minimize the first addend of the above variance and also the total variance asymptotically. Those weights are in line with the standard weights from Zhang [2006] in the univariate setting and the following minimization is analogous as well.

Minimization with side conditions (C1) and (C2) yields for an arbitrary constant c

$$
\begin{aligned}
&\frac{\partial}{\partial\alpha_j}\left(c\sum_{i=1}^{M_N}\frac{\alpha_{i,M_N}^2}{i^2}+\lambda_1\left(\sum_{i=1}^{M_N}\alpha_{i,M_N}-1\right)+\lambda_2\left(\sum_{i=1}^{M_N}\frac{\alpha_{i,M_N}}{i}\right)\right)=0\\
\Leftrightarrow\quad & 2c\frac{\alpha_{j,M_N}}{j^2}+\lambda_1+\frac{\lambda_2}{j}=0\\
\Leftrightarrow\quad & \alpha_{j,M_N}=-\frac{1}{2c}\left(j^2\lambda_1+\lambda_2 j\right)\ .
\end{aligned}
$$

Since

$$1=\sum_j\alpha_j=-(1/2c)\left(\lambda_1\sum_j j^2+\lambda_2\sum_j j\right)$$

and

$$0=\sum_j(\alpha_{j,M_N}/j)=-(1/2c)\left(\lambda_1\sum_j j+\lambda_2 M_N\right)\ ,$$

we obtain the result

$$\lambda_1=\frac{-24c}{M_N^3-M_N}\ ,\ \lambda_2=\frac{12c}{(M_N-1)M_N}\ ,\ \alpha_{j,M_N}^{opt}=\frac{12j^2-6jM_N-6j}{(M_N^3-M_N)}\ .$$

Moreover,

$$
\begin{aligned}
&\sum_{j=1}^{N}\sum_{i=1}^{M_N-1}\eta_X^2\eta_Y^2\left(\frac{\alpha_{i,M_N}^{opt}\alpha_{i+1,M_N}^{opt}}{i(i+1)}-\left(\frac{\alpha_{i,M_N}^{opt}}{i}\right)^2\right)\\
&=\sum_{j=1}^{N}\sum_{i=1}^{M_N-1}\eta_X^2\eta_Y^2\frac{12}{M_N^3}\frac{12i-6M_N-6}{M_N^3-M_N}=\mathcal{O}\left(NM_N^{-4}\right)=\scriptstyle{\mathcal{O}}\left(NM_N^{-3}\right).
\end{aligned}
$$

Inserting the noise-optimal weights (4.14), we can apply the central limit theorem for strong mixing triangular arrays 1.9 to

$$\sqrt{\frac{M_N^3}{N}}\sum_{j=1}^{N}\frac{\alpha_{i,M_N}^{opt}}{i}\sum_{i=1}^{M_N\wedge j}\left(\epsilon_{g_j}^X\epsilon_{\lambda_{j-i+1}}^Y+\epsilon_{l_{j-i+1}}^X\epsilon_{\gamma_j}^Y\right)\ .$$

The sequence of variances with the chosen weights according to (4.14)

$$\mathbb{V}\mathrm{ar}\left(\sqrt{\frac{M_N^3}{N}}\sum_{j=1}^{N}\frac{\alpha_{i,M_N}^{opt}}{i}\sum_{i=1}^{M_N\wedge j}\left(\epsilon_{g_j}^X\epsilon_{\lambda_{j-i+1}}^Y+\epsilon_{l_{j-i+1}}^X\epsilon_{\gamma_j}^Y\right)\right)$$

$$
\begin{aligned}
&= \frac{M_N^3}{N} \sum_{j=0}^{N} \sum_{i=1}^{M_N \wedge (j+1)} \left(\frac{\alpha_{i,M_N}^{opt}}{i} \right)^2 \eta_X^2 \eta_Y^2 (3 - \mathbb{1}_{\{g_j \neg_{j+1}\,,\, \gamma_j \neq \gamma_{j+1}\}}) + \mathcal{o}(1) \\
&\longrightarrow 36 \eta_X^2 \eta_Y^2 - 12(1-\kappa_{\mathrm{n}}) \eta_X^2 \eta_Y^2
\end{aligned}
$$

converges to $\mathbf{AVAR}_{\mathrm{n}}$ on the Assumption 4.2.
Since the inner sums are 2-dependent and hence in particular ϕ-mixing, it suffices to prove the Lyapunov condition (LY) to apply Theorem 1.9. For $\delta = 2$ we have the following sum of fourth moments

$$
\sum_{j=1}^{N} \frac{M_N^6}{N^2} \left(\sum_{i=1}^{M_N \wedge j} \left(\frac{\alpha_{i,M_N}^{opt}}{i} \right)^4 2\mathbb{E}\left[\left(\epsilon_{t_1}^X \epsilon_{\tau_1}^Y \right)^4 \right] + 4(\eta_X^2 \eta_Y^2)^2 \left(\sum_{i=1}^{M_N \wedge j} \frac{\left(\alpha_{i,M_N}^{opt}\right)^2}{i^2} \right)^2 \right) = \mathcal{O}\left(N^{-1}\right)
$$

which is a null sequence. The first addend is $\mathcal{o}(N^{-1})$ because the inner sum is of order M_N^{-7} and the second addend is of order $\mathcal{O}(N^{-1})$ because the inner sum is of order M_N^{-6}. This completes the proof of the proposition. □

Next, we consider the remaining addends of the error induced by microstructure frictions and insert the weights (4.14):

Proposition 4.3.2. *On the Assumptions 4.1, 4.2 and 3, the following weak convergence to a centred normal distribution holds true:*

$$
\sqrt{M_N} \left(\sum_{i=1}^{M_N} \frac{\alpha_{i,M_N}^{opt}}{i} \left(\sum_{j=1}^{i-1} \epsilon_{g_j}^X \epsilon_{\gamma_j}^Y + \sum_{j=N-i+1}^{N} \epsilon_{l_j}^X \epsilon_{\lambda_j}^Y \right) \right) \rightsquigarrow \mathbf{N}\left(0, \frac{12}{5} \eta_X^2 \eta_Y^2 \right) . \tag{4.18}
$$

This convergence also holds conditionally on the paths of the efficient processes.

Proof.

$$
\sum_{i=1}^{M_N} \frac{\alpha_{i,M_N}^{opt}}{i} \left(\sum_{j=1}^{i-1} \epsilon_{g_j}^X \epsilon_{\gamma_j}^Y + \sum_{j=N-i+1}^{N} \epsilon_{l_j}^X \epsilon_{\lambda_j}^Y \right) = \sum_{j=1}^{M_N-1} \left(\epsilon_{g_j}^X \epsilon_{\gamma_j}^Y + \epsilon_{l_{N-j}}^X \epsilon_{\lambda_{N-j}}^Y \right) \sum_{i=j+1}^{M_N} \frac{\alpha_{i,M_N}^{opt}}{i}
$$

Both addends are uncorrelated and $\sqrt{M_N} \sum_{j=1}^{M_N-1} \epsilon_{g_j}^X \epsilon_{\gamma_j}^Y \sum_{i=j+1}^{M_N} \alpha_{i,M_N}^{opt}/i$ is the endpoint of a discrete centred martingale with respect to the filtration

$$
\overline{\mathcal{A}}_j^N := \sigma\left(\epsilon_{t_k}^X | t_k \le g_j \,,\, X_{t_k} | 0 \le k \le n \right) \vee \sigma\left(\epsilon_{\tau_k}^Y | \tau_k \le \gamma_j \,,\, Y_{\tau_k} | 0 \le k \le m \right) .
$$

Namely $g_j = g_{j-1} \Rightarrow \gamma_j > \gamma_{j-1}$ and analogously $\gamma_j = \gamma_{j-1} \Rightarrow g_j > g_{j-1}$ and hence

$$
\begin{aligned}
\mathbb{E}\left[\sqrt{M_N} \epsilon_{g_l}^X \epsilon_{\gamma_l}^Y \sum_{i=l+1}^{M_N} \frac{\alpha_{i,M_N}^{opt}}{i} \Big| \overline{\mathcal{A}}_{l-1}^N \right] &= \sqrt{M_N} \Big(\mathbb{1}_{\{g_l = g_{l-1}\}} \mathbb{E}\left[\epsilon_{\gamma_l}^Y\right] \epsilon_{g_{l-1}}^X \\
&+ \mathbb{1}_{\{\gamma_l = \gamma_{l-1}\}} \mathbb{E}\left[\epsilon_{g_l}^X\right] \epsilon_{\gamma_{l-1}}^Y + \mathbb{1}_{\{g_l \neq g_{l-1}\,,\, \gamma_l \neq \gamma_{l-1}\}} \mathbb{E}\left[\epsilon_{g_l}^X\right] \mathbb{E}\left[\epsilon_{\gamma_l}^Y\right] \Big) = 0 \,.
\end{aligned}
$$

The central limit Theorem for martingales 1.7 will be applied. The conditional Lindeberg condition (C–LB) can be proved by verifying the stronger conditional Lyapunov condition (C–LY) with $\delta = 2$ (cf. Corollary 1.3.2 and 1.3.3):

$$\mathbb{E}\left[\left(\sqrt{M_N}\sum_{j=1}^{M_N-1}\epsilon_{g_j}^X\epsilon_{\gamma_j}^Y\sum_{i=j+1}^{M_N}\frac{\alpha_{i,M_N}^{opt}}{i}\right)^4\Big|\overline{\mathcal{A}}_{l-1}^N\right] = \sum_{j=1}^{M_N-1}\left(\sqrt{M_N}\sum_{i=j+1}^{M_N}\frac{\alpha_{i,M_N}^{opt}}{i}\right)^4$$
$$\times\left(\mathbb{E}\left[\left(\epsilon_{g_j}^X\epsilon_{\gamma_j}^Y\right)^4\right]\mathbb{1}_{\{g_l\neq g_{l-1}\,,\,\gamma_l\neq\gamma_{l-1}\}}\right.$$
$$\left.+\mathbb{E}\left[\left(\epsilon_{g_j}^X\right)^4\right]\left(\epsilon_{\gamma_j}^Y\right)^4\mathbb{1}_{\{\gamma_l=\gamma_{l-1}\}}+\mathbb{E}\left[\left(\epsilon_{\gamma_j}^Y\right)^4\right]\left(\epsilon_{g_j}^X\right)^4\mathbb{1}_{\{g_l=g_{l-1}\}}\right)$$
$$\xrightarrow{p} 0\ .$$

The stochastic convergence to zero holds because of $\sum_j\left(\sum_{i=j+1}^{M_N}(\alpha_{i,M_N}^{opt}/i)\right)^4 = (72/35)$ $M_n^{-3}+\mathcal{O}(M_N^{-3})$ and the law of large numbers. The asymptotic conditional variance equals

$$M_N\sum_{j=1}^{M_N-1}\mathbb{V}\mathrm{ar}\left(\epsilon_{g_j}^X\epsilon_{\gamma_j}^Y\sum_{i=j+1}^{M_N}\frac{\alpha_{i,M_N}^{opt}}{i}\Big|\mathcal{A}_{N,j-1}\right) = M_N\sum_{j=1}^{M_N-1}\left(\sum_{i=j+1}^{M_N}\frac{\alpha_{i,M_N}^{opt}}{i}\right)^2$$
$$\times\left(\eta_X^2\eta_Y^2\mathbb{1}_{\{g_j\neq g_{j-1}\,,\,\gamma_j\neq\gamma_{j-1}\}}+\left(\epsilon_{g_j}^X\right)^2\mathbb{1}_{\{g_j=g_{j-1}\}}\eta_Y^2+\left(\epsilon_{\gamma_j}^Y\right)^2\mathbb{1}_{\{\gamma_j=\gamma_{j-1}\}}\eta_X^2\right)$$
$$\xrightarrow{p}\frac{6}{5}\eta_X^2\eta_Y^2\ .$$

We have used the formula $\sum_{j=1}^{M_N-1}\left(\sum_{i=j+1}M_N(\alpha_{i,M_N}^{opt}/i)\right)^2 = (6/5)M_N^{-1}+\mathcal{O}(M_N^{-1})$.

The second addend $\sum_{j=1}^{M_N-1}\left(\epsilon_{l_{N-j}}^X\epsilon_{\lambda_{N-j}}^Y\right)\sum_{i=j+1}^{M_N}\frac{\alpha_{i,M_N}^{opt}}{i}$ is the endpoint of a discrete centred martingale with respect to the filtration $\underline{\mathcal{A}}_{N-j}^N$ where

$$\underline{\mathcal{A}}_j^N := \sigma\left(\epsilon_{t_k}^X|t_k\le l_j\,,\,X_{t_k}|0\le k\le n\right)\vee\sigma\left(\epsilon_{\tau_k}^Y|\tau_k\le\lambda_j\,,\,Y_{\tau_k}|0\le k\le m\right)$$

and converges multiplied with $\sqrt{M_N}$ to a centred mixed Gaussian distribution with same asymptotic variance $(6/5)\eta_X^2\eta_Y^2$. The proof is analogous as for the first addend when we leave out the indicator functions because the minima are all different observation times. The proposition is concluded using Theorem 1.7. □

4.3.2 Asymptotic discretization error of the one-scale subsampling estimator and the generalized multiscale estimator

Proposition 4.3.3. *On the Assumptions 1, 4.1 and 4.3, the discretization error of the one-scale subsampling estimator with subsampling frequency i_N converges stably in law* *to a centred mixed normal limit as $i_N \to \infty, N \to \infty, i_N/N^\alpha \to 0$ for every $\alpha > 2/3$:*

$$\sqrt{\frac{N}{i_N}}\left(\sum_{j=i}^{N}\left(X_{g_j}-X_{l_{j-i+1}}\right)\left(Y_{\gamma_j}-Y_{\lambda_{j-i+1}}\right)-\langle X,Y\rangle_T\right)\overset{st}{\rightsquigarrow}\mathbf{N}\left(0,\mathbf{AVAR}_{dis,sub}\right)\ ,$$

with asymptotic variance

$$\mathbf{AVAR}_{dis,sub}=\frac{2}{3}T\int_0^T G'(t)(\sigma_t^X\sigma_t^Y)^2(1+\rho_t^2)\,dt\ . \tag{4.19}$$

Proposition 4.3.4. *On the Assumptions 1,4.1 and 4.3, the discretization error of the generalized multiscale estimator with the noise-optimal weights given in* (4.14) *converges with rate $\sqrt{N/M_N}$ stably in law to a centred mixed Gaussian limit as $M_N \to \infty, N \to \infty, M_N/N^\alpha \to 0$ for every $\alpha > 2/3$:*

$$\sqrt{\frac{N}{M_N}}\left(\sum_{i=1}^{M_N}\frac{\alpha_{i,M_N}^{opt}}{i}\sum_{j=i}^{N}\left(X_{g_j}-X_{l_{j-i+1}}\right)\left(Y_{\gamma_j}-Y_{\lambda_{j-i+1}}\right)-\langle X,Y\rangle_T\right)\overset{st}{\rightsquigarrow}\mathbf{N}\left(0,\mathbf{AVAR}_{dis}^{multi}\right)$$

with asymptotic variance

$$\mathbf{AVAR}_{dis}^{multi}=\frac{26}{35}T\int_0^T G'(t)(\sigma_t^X\sigma_t^Y)^2(1+\rho_t^2)\,dt\ . \tag{4.20}$$

In the following, it is proved that the errors due to the lack of synchronicity for the one-scale subsampling estimator and the generalized multiscale estimator are asymptotically negligible. Therefore, Propositions 4.3.3 and 4.3.4 above are implied by the corresponding limit theorems for the discretization errors of the closest synchronous approximation given as Propositions 4.3.5 and 4.3.6 in the next paragraph.
Note that M_N is chosen of order $\sqrt{N}$ in Theorem 4.1, but the condition $M_N = \mathcal{O}\left(N^{2/3}\right)$ is needed here as a regularity assumption.

Discretization error of the closest synchronous approximation

Proposition 4.3.5. *Suppose that Assumptions 1, 4.1 and 4.3 hold true. As $i_N \to \infty, N \to \infty, i_N/N^\alpha \to 0$ for every $\alpha > 2/3$, the discretization error of the closest synchronous approximation for the one-scale subsampling estimator with subsampling frequency i_N converges stably in law to a centred mixed normal limit:*

$$\sqrt{\frac{N}{i_N}}\left(\sum_{j=i}^{N}\left(X_{T_j}-X_{T_{j-i}}\right)\left(Y_{T_j}-Y_{T_{j-i}}\right)-\langle X,Y\rangle_T\right)\overset{st}{\rightsquigarrow}\mathbf{N}\left(0,\mathbf{AVAR}_{syn,sub}\right)\ ,$$

with asymptotic variance

$$\mathbf{AVAR}_{syn,sub}=\frac{2}{3}T\int_0^T G'(t)(\sigma_t^X\sigma_t^Y)^2(1+\rho_t^2)\,dt\ .$$

Proposition 4.3.6. *Let Assumptions 1,4.1 and 4.3 hold true. For the discretization error of the closest synchronous approximation of the generalized multiscale estimator with the noise-optimal weights given in* (4.14)*, the following stable central limit theorem holds true as* $M_N \to \infty, N \to \infty, M_N/N^\alpha \to 0$ *for every* $\alpha > 2/3$*:*

$$\sqrt{\frac{N}{M_N}}\left(\sum_{i=1}^{M_N}\frac{\alpha_{i,M_N}^{opt}}{i}\sum_{j=i}^{N}\left(X_{T_j}-X_{T_{j-i}}\right)\left(Y_{T_j}-Y_{T_{j-i}}\right)-\langle X,Y\rangle_T\right)\overset{st}{\rightsquigarrow}\mathbf{N}\left(0,\mathbf{AVAR}_{syn}^{multi}\right)$$

with asymptotic variance

$$\mathbf{AVAR}_{syn}^{multi}=\frac{26}{35}T\int_0^T G'(t)(\sigma_t^X\sigma_t^Y)^2(1+\rho_t^2)\,dt\ .$$

Note that it suffices to prove the foregoing limit theorems for the zero-drift case. Since our limit theorems are stable, asymptotic mixed normality is assured to hold for the general setting on Assumption 1. This has been emphasized in Subsection 1.1.2 and at the end of Section 1.2. Denote $L_t=\int_0^t\sigma_s^X dW_s^X$ and $M_t=\int_0^t\sigma_s^Y dW_s^Y$ the continuous martingales that represent the efficient processes under the equivalent martingale measure. As before in Section 3.3 the asymptotic mixed normality is implied as marginal distribution of a limiting time-changed Brownian motion which is proven to be the weak limit of the process corresponding to the discretization error.
We start with the discretization error of the closest synchronous approximation of a one-scale subsampling estimator. Recall the illustration of the discretization error in (4.11).

Proposition 4.3.7. *On the same assumptions as in Proposition 4.3.5, the continuous martingale*

$$\begin{aligned}\mathfrak{D}_t^N:=\sqrt{\frac{N}{i_NT}}\Bigg[&\sum_{T_k\le t}\left(\Delta L_{T_k}+\int_{T_k}^t\sigma_s^X dW_s^X\right)\left(\sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)\Delta M_{T_{k-l}}\right)\\&+\sum_{T_k\le t}\left(\Delta M_{T_k}+\int_{T_k}^t\sigma_s^Y dW_s^Y\right)\left(\sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)\Delta L_{T_{k-l}}\right)\Bigg]\end{aligned}$$

for $t\in[0,T]$*, where* $\Delta\,\cdot_{T_k}=\cdot_{T_k}-\cdot_{T_{k-1}}$ *is the backward difference operator, converges stably in law as* $N\to\infty, i_N\to\infty, i_N/N\to 0$ *to a limiting time-changed Brownian motion*

$$\mathfrak{D}_t^N\overset{st}{\rightsquigarrow}\int_0^t\sqrt{v_{\mathfrak{D}_s}}\,d\mathfrak{W}_s^\perp\ ,$$

where $\mathfrak{W}^\perp$ *is independent of* $\mathcal{F}$ *and*

$$v_{\mathfrak{D}_s}=\frac{2}{3}G'(s)(\sigma_s^X\sigma_s^Y)^2(1+\rho_s^2)\ .$$

Proof of Proposition 4.3.7:
In the successional proof the subscript of the subsampling frequency is omitted and C denotes a generic constant and $\delta_N = \sup_{i\in\{1,\dots,N\}}(T_i - T_{i-1})$.
Jacod's Theorem 1.6 is applied in the manner of the proof of Proposition 3.3.2. Note that the conditions (1.3a), (1.3b) and the convergence of the quadratic variation process correspond to the conditions of the discrete-time version of the theorem, except we spare to prove a Lindeberg-type condition when working with the complemented continuous-time martingales.
Calculating the quadratic variation of $\mathfrak{D}_t^N$ yields

$$
\begin{aligned}
\langle \mathfrak{D}^N\rangle_t = \frac{N}{iT}\Bigg[&\sum_{T_k\le t}\left(\Delta\langle L\rangle_{T_k}\left(\sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)\Delta M_{T_{k-l}}\right)^2 + \Delta\langle M\rangle_{T_k}\left(\sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)\Delta L_{T_{k-l}}\right)^2\right)\\
&+2\sum_{T_k\le t}\Delta\langle L,M\rangle_{T_k}\left(\sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)\Delta L_{T_{k-l}}\right)\left(\sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)\Delta M_{T_{k-l}}\right)\Bigg] + \mathcal{O}_p(1)\\
\underset{\text{Lemma 4.3.8}}{=} \frac{N}{iT}&\Bigg[\sum_{T_k\le t}\Delta\langle L\rangle_{T_k}\sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)^2\left(\Delta M_{T_{k-l}}\right)^2 + \sum_{T_k\le t}\Delta\langle M\rangle_{T_k}\\
\times\sum_{l=1}^{i\wedge k}&\left(1-\frac{l}{i}\right)^2\left(\Delta L_{T_{k-l}}\right)^2 + 2\sum_{T_k\le t}\Delta\langle L,M\rangle_{T_k}\sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)^2\left(\Delta L_{T_{k-l}}\right)^2\Bigg] + \mathcal{O}_p(1)\\
\underset{\text{Lemma 4.3.9}}{=} \frac{N}{iT}&\Bigg[\sum_{T_k\le t}\int_{T_{k-1}}^{T_k}(\sigma_s^X)^2ds\left(\sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)^2\int_{T_{k-l-1}}^{T_{k-l}}(\sigma_s^Y)^2ds\right)\\
&+\sum_{T_k\le t}\int_{T_{k-1}}^{T_k}(\sigma_s^Y)^2ds\left(\sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)^2\int_{T_{k-l-1}}^{T_{k-l}}(\sigma_s^X)^2ds\right)\\
&+2\sum_{T_k\le t}\int_{T_{k-1}}^{T_k}\rho_s\sigma_s^X\sigma_s^Y\,ds\left(\sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)^2\int_{T_{k-l-1}}^{T_{k-l}}\rho_s\sigma_s^X\sigma_s^Y\,ds\right)\Bigg] + \mathcal{O}_p(1)\\
\underset{\text{Lemma 4.3.10}}{=} \frac{N}{iT}&\sum_{T_k\le t}2(1+\rho_{T_{k-1}}^2)(\sigma_{T_{k-1}}^X\sigma_{T_{k-1}}^Y)^2\,(\Delta T_k)^2\sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)^2 + \mathcal{O}_p(1)\\
= \sum_{T_k\le t}&\frac{2}{3}\frac{G^N(T_k)-G^N(T_{k-1})}{T_k-T_{k-1}}\left(\rho_{T_{k-1}}^2+1\right)\left(\sigma_{T_{k-1}}^X\sigma_{T_{k-1}}^Y\right)^2\Delta T_k + \mathcal{O}_p(1)\\
\xrightarrow{p}\ &\frac{2}{3}\int_0^t(1+\rho_s^2)(\sigma_s^X\sigma_s^Y)^2G'(s)\,ds\quad .
\end{aligned}
$$

In the first step cross terms of the inner sums have been neglected. In the following step the squared increments of L and M and the increments of the product $L\cdot M$ in these inner sums are substituted by the increments of the quadratic (co-)variation processes. Next, on each block $(T_{k-1},\dots,T_{k-i\vee 0})$ the increments of the form $\int_{T_{k-l-1}}^{T_{k-l}}f(t)dt$ with continuous functions f for $l=1,\dots,k\wedge i$ are approximated by $\Delta T_k f(T_{k-1})$. The error induced by the first two steps is shown to be asymptotically negligible in Lemmas 4.3.8 and 4.3.9 with similar methods as already needed for the proof of Theorem 3.1. The crucial blockwise approximation that leads to the simple closed-form expression for the asymptotic quadratic variation $\langle \mathfrak{D}^N \rangle$ is treated in Lemma 4.3.10. Here the concept of a time-changed quadratic variation of time comes into play and is needed to prove this approximation. It has been further used that $1/i\sum_{l=1}^{i}(1-(l/i))^2 = 1/3 + \mathcal{o}(1)$ and the convergence in probability follows by Assumption 4.3 and the convergence of the Riemann sums to the integral.

Lemma 4.3.8. *On the same assumptions as in Proposition 4.3.5, it holds true that*

$$2\sum_{T_k\le t}\Delta\langle L\rangle_{T_k}\left(\sum_{r=2}^{i\wedge k}\sum_{q=1}^{(r-1)\wedge(k-1)}\left(1-\frac{r}{i}\right)\left(1-\frac{q}{i}\right)\Delta M_{T_{k-r}}\Delta M_{T_{k-q}}\right)=\mathcal{o}_p\left(iN^{-1}\right)\,,$$

$$2\sum_{T_k\le t}\Delta\langle M\rangle_{T_k}\left(\sum_{r=2}^{i\wedge k}\sum_{q=1}^{(r-1)\wedge(k-1)}\left(1-\frac{r}{i}\right)\left(1-\frac{q}{i}\right)\Delta L_{T_{k-r}}\Delta L_{T_{k-q}}\right)=\mathcal{o}_p\left(iN^{-1}\right)\,,$$

$$2\sum_{T_k\le t}\Delta\langle L,M\rangle_{T_k}\left(\sum_{r=1}^{i\wedge k}\sum_{q\ne r\in\{1,\dots,i\wedge k\}}\left(1-\frac{r}{i}\right)\left(1-\frac{q}{i}\right)\Delta M_{T_{k-r}}\Delta L_{T_{k-q}}\right)=\mathcal{o}_p\left(iN^{-1}\right)\,.$$

Proof. All three terms have an expectation equal to zero. The asymptotic orders of the three terms are deduced following the same principles and we restrict us to the proof of the third equation. The left-hand side can be written

$$2\sum_{T_k\le t}\Delta\langle L,M\rangle_{T_k}\xi_k$$

with centred i-dependent random variables

$$\xi_k=\sum_{r=1}^{i\wedge k}\sum_{q\ne r\in\{1,\dots,i\wedge k\}}\left(1-\frac{r}{i}\right)\left(1-\frac{q}{i}\right)\Delta M_{T_{k-r}}\Delta L_{T_{k-q}}\,.$$

Since Brownian increments over disjoint time intervals are independent, applying Itô isometry the variances $\mathbb{V}\mathrm{ar}(\xi_k), k=1,\dots,N$ are bounded by

$$\mathbb{V}\mathrm{ar}(\xi_k)=\sum_{r=1}^{i\wedge k}\sum_{q\ne r\in\{1,\dots,i\wedge k\}}\left(1-\frac{r}{i}\right)^2\left(1-\frac{q}{i}\right)^2\mathbb{E}\left[\int_{T_{k-r-1}}^{T_{k-r}}(\sigma_s^X)^2ds\int_{T_{k-q-1}}^{T_{k-q}}(\sigma_s^Y)^2ds\right.$$
$$\left.+\int_{T_{k-r-1}}^{T_{k-r}}\rho_s\sigma_s^X\sigma_s^Y\,ds\int_{T_{k-q-1}}^{T_{k-q}}\rho_s\sigma_s^X\sigma_s^Y\,ds\right]$$

$$\leq C\delta_N^2 \sum_{r=1}^{i\wedge k} \sum_{q\neq r\in\{1,\ldots,i\wedge k\}} \left(1-\frac{r}{i}\right)^2 \left(1-\frac{q}{i}\right)^2 = \mathcal{O}\left(i^2\delta_N^2\right) \ .$$

Thus, the second moment of the sum above is bounded by

$$\mathbb{E}\left[\left(2\sum_{T_k\leq t} \Delta\langle L,M\rangle_{T_k}\xi_k\right)^2\right] \leq C\delta_N^2\mathbb{E}\left[\left(\sum_{k=1}^N \xi_k\right)^2\right] = \mathcal{O}\left(i^3\delta_N^3\right) \ .$$

We derive that the term is of order $\mathcal{O}_p(iN^{-1})$ as long as $\delta_N = \mathcal{o}\left(i^{-1/3}N^{-2/3}\right)$ which is assured by the condition $i = \mathcal{O}\left(N^{2/3}\right)$ and Assumption 2(b). □

Lemma 4.3.9. *On the same assumptions as in Proposition 4.3.5, it holds true that*

$$\frac{N}{iT}\sum_{T_k\leq t}\Delta\langle L\rangle_{T_k}\left(\sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)^2\left(\Delta M_{T_{k-l}}\right)^2 - \sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)^2\int_{T_{k-l-1}}^{T_{k-l}}(\sigma_s^Y)^2ds\right) = \mathcal{o}_p(1) \ ,$$

$$\frac{N}{iT}\sum_{T_k\leq t}\Delta\langle M\rangle_{T_k}\left(\sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)^2\left(\Delta L_{T_{k-l}}\right)^2 - \sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)^2\int_{T_{k-l-1}}^{T_{k-l}}(\sigma_s^X)^2ds\right) = \mathcal{o}_p(1) \ ,$$

$$\frac{N}{iT}\sum_{T_k\leq t}2\Delta\langle L,M\rangle_{T_k}\left(\sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)^2\left(\Delta L_{T_{k-l}}\Delta M_{T_{k-l}} - \int_{T_{k-l-1}}^{T_{k-l}}\rho_s\sigma_s^Y\sigma_s^X ds\right)\right) = \mathcal{o}_p(1) \ .$$

Proof. The left-hand sides are centred due to Itô isometry. The three equations can be proved analogously and we restrict ourselves to prove the first one. Denote

$$\psi_k = \sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)^2\left(\Delta M_{T_{k-l}}\right)^2 - \sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)^2\int_{T_{k-l-1}}^{T_{k-l}}(\sigma_s^Y)^2ds \ \ , k\in\{1,\ldots,N\},$$

that are i-dependent centred random variables with finite variances

$$\begin{aligned}\mathbb{E}\psi_k^2 &= \sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)^4\mathbb{E}\left[\left(\left(\Delta M_{T_{k-l}}\right)^2 - \Delta\langle M\rangle_{T_{k-l}}\right)^2\right]\\ &\leq C\delta_N^2\sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)^4 = \mathcal{O}\left(i\delta_N^2\right) \ .\end{aligned}$$

It has been used that the cross terms are uncorrelated by Itô isometry. For fixed $k\,,(N-i) > k > (i+1)$, ψ_k is correlated with $\psi_r\,,\ r\in\{k-i,\ldots,k+i\}$, and the sum of all correlations ($r \neq k$) is bounded by

$$\sum_{r=k-i}^{k-1} \mathbb{E}\left[\psi_k\psi_r\right] + \sum_{r=k+1}^{k+i} \mathbb{E}\left[\psi_k\psi_r\right] \leq \sum_{r=k-i}^{k-1} \left(\mathbb{E}\psi_k^2\mathbb{E}\psi_r^2\right)^{\frac{1}{2}} + \sum_{r=k+1}^{k+i} \left(\mathbb{E}\psi_k^2\mathbb{E}\psi_r^2\right)^{\frac{1}{2}} = \mathcal{O}\left(i^2\delta_N^2\right) \ .$$

The asymptotic order of the variances $\mathbb{E}\psi_k^2$ calculated above and the Cauchy-Schwarz inequality has been used. We conclude the order of the second moment

$$\begin{aligned}\mathbb{E}\left[\left(\frac{N}{iT}\sum_{T_k\leq t}\Delta\langle L\rangle_{T_k}\psi_k\right)^2\right] &= \frac{N^2}{i^2T^2}\sum_{T_k\leq t}\Bigg(\mathbb{E}\left[(\Delta\langle L\rangle_{T_k})^2\right]\mathbb{E}\psi_k^2 \\ &\quad + \sum_{r=k-i\vee 0}^{k-1} 2\mathbb{E}\left[\Delta\langle L\rangle_{T_k}\Delta\langle L\rangle_{T_r}\psi_k\psi_r\right]\Bigg) \\ &\leq C\frac{N^2\delta_N^2}{i^2T^2}\sum_{T_k\leq t}\left(\mathbb{V}\mathrm{ar}(\psi_k) + \sum_{r=k-i\vee 0}^{k-1}\mathbb{E}\left[\psi_k\psi_r\right]\right) = \mathcal{O}\left(\delta_N\right) \ .\end{aligned}$$

Hence, the term converges to zero in probability. □

Lemma 4.3.10. *On the same assumptions as in Proposition 4.3.5, it holds true that the terms*

$$\frac{N}{iT}\sum_{T_k\leq t}\left(\int_{T_{k-1}}^{T_k}(\sigma_s^X)^2ds\left(\sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)^2\left(\int_{T_{k-l-1}}^{T_{k-l}}(\sigma_s^Y)^2\,ds - (\sigma^X_{T_{k-1}}\sigma^Y_{T_{k-1}}\Delta T_k)^2\right)\right)\right) \ ,$$

$$\frac{N}{iT}\sum_{T_k\leq t}\left(\int_{T_{k-1}}^{T_k}(\sigma_s^Y)^2ds\left(\sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)^2\left(\int_{T_{k-l-1}}^{T_{k-l}}(\sigma_s^X)^2\,ds - (\sigma^X_{T_{k-1}}\sigma^Y_{T_{k-1}}\Delta T_k)^2\right)\right)\right) \ ,$$

$$\begin{aligned}\frac{N}{iT}\sum_{T_k\leq t}\Bigg(\int_{T_{k-1}}^{T_k}\rho_s\sigma_s^X\sigma_s^Y\,ds&\left(\sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)^2\int_{T_{k-l-1}}^{T_{k-l}}\rho_s\sigma_s^X\sigma_s^Y\,ds\right) \\ &-(\rho_{T_{k-1}}\sigma^X_{T_{k-1}}\sigma^Y_{T_{k-1}}\Delta T_k)^2\sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)^2\Bigg) \ ,\end{aligned}$$

converge to zero in probability.

Proof. The proof that the approximation errors of the type above converge to zero in probability is based on the concept of a time-change in the asymptotic quadratic variation of refresh times from Assumption 3.1 which has been presented and discussed in the last section as part of the preparations to establish a central limit theorem. We will suppose without loss of generality that the sampling design of the closest synchronous approximation satisfies (4.13).

First, an application of the mean value theorem yields

$$\begin{aligned}
&\frac{N}{iT}\sum_{T_k\le t}\int_{T_{k-1}}^{T_k}(\sigma_s^X)^2ds\left(\sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)^2\int_{T_{k-l-1}}^{T_{k-l}}(\sigma_s^Y)^2ds\right)\\
&\quad=\frac{N}{iT}\sum_{T_k\le t}(\sigma_{\zeta_k}^X)^2\Delta T_k\left(\sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)^2(\sigma_{\zeta_{k-l}^*}^Y)^2\Delta T_{k-l}\right)
\end{aligned}$$

with $\zeta_k\in[T_{k-1},T_k]$, $\zeta_q^*\in[T_{q-1},T_q]$. Since the volatility processes σ^X,σ^Y are uniformly continuous on $[0,T]$ by Assumption 1

$$\sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)^2\left|(\sigma_{\zeta_{k-l}^*}^Y)^2-(\sigma_{T_{k-1}}^Y)^2\right|\Delta T_{k-l}\le i\delta_N\sup_{|t-s|\le i\delta_N}\left|(\sigma_t^Y)^2-(\sigma_s^Y)^2\right|=\mathcal{O}_{a.\,s.}(i\delta_N)\ ,$$

$$\sum_{T_k\le t}\left|(\sigma_{\zeta_k}^X)^2-(\sigma_{T_{k-1}}^X)^2\right|(\sigma_{T_{k-1}}^Y)^2\Delta T_k\sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)^2\Delta T_{k-l}\le i\delta_N\sup_{|t-s|\le\delta_N}\left|(\sigma_t^Y)^2-(\sigma_s^Y)^2\right|$$

which is $\mathcal{O}_{a.\,s.}(i\delta_N)$. The asymptotic orders even hold almost surely which is represented by $\mathcal{O}_{a.\,s.}$. Since ρ is as well uniformly continuous analogous conclusions hold true for the two other terms.
Using the Cauchy-Schwarz inequality and (4.13), we obtain

$$\begin{aligned}
&\frac{N}{iT}\sum_{T_k\le t}(\sigma_{T_{k-1}}^X\sigma_{T_{k-1}}^Y)^2\Delta T_k\left(\sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)^2\left|\Delta T_{k-l}-\frac{T}{N}\right|\right)\\
&\quad\le\frac{N}{iT}\sup_{s\in[0,t]}\left(\sigma_t^X\sigma_t^Y\right)^2\sum_{l=1}^{i}\sum_{j=1}^{N-l}\left|\left(\Delta T_j-\frac{T}{N}\right)\Delta T_{j+l}\right|\\
&\quad\le\frac{N}{iT}C\left(\sum_{j=1}^{N}\left(T_{(j+i)\vee N}-T_j\right)^2\sum_{j=1}^{N}\left(\Delta T_j-\frac{T}{N}\right)^2\right)^{1/2}=\mathcal{O}_{a.\,s.}(1)\ .
\end{aligned}$$

Furthermore,

$$\begin{aligned}
&\frac{N}{iT}\sum_{T_k\le t}(\sigma_{T_{k-1}}^X\sigma_{T_{k-1}}^Y)^2\Delta T_k\left|\Delta T_k-\frac{T}{N}\right|\sum_{l=1}^{i\wedge k}\left(1-\frac{l}{i}\right)^2\\
&\quad\le\frac{N}{T}C\left(\sum_{j=1}^{N}(\Delta T_j)^2\sum_{j=1}^{N}\left(\Delta T_j-\frac{T}{N}\right)^2\right)^{1/2}=\mathcal{O}_{a.\,s.}(1)
\end{aligned}$$

holds, also almost surely, due to (4.13) and the Cauchy-Schwarz inequality. The last two approximations hold analogously for the two other terms. The four preceding approximation errors that were shown to converge to zero almost surely imply the statement of the lemma. □

We proceed proving that the quadratic covariations $\langle \mathfrak{D}^N, L\rangle_t$ and $\langle \mathfrak{D}^N, M\rangle_t$ converge to zero in probability for all $t \in [0,T]$.

$$\langle \mathfrak{D}^N, L\rangle_t = \sqrt{\frac{N}{iT}} \sum_{T_k \le t} \left(\Delta \langle L\rangle_{T_k} \left(\sum_{l=1}^{i\wedge k} \left(1 - \frac{l}{i}\right) \Delta M_{T_{k-l}} \right) + \Delta \langle L, M\rangle_{T_k} \left(\sum_{l=1}^{i\wedge k} \left(1 - \frac{l}{i}\right) \Delta L_{T_{k-l}} \right) \right)$$

has an expectation equal to zero for all $t \in [0,T]$ and the second moment is bounded above by

$$\frac{N}{iT} \mathbb{E}\left[\left(\sum_{T_k \le t} \left(\Delta \langle L\rangle_{T_k} \left(\sum_{l=1}^{i\wedge k} \left(1 - \frac{l}{i}\right) \Delta M_{T_{k-l}} \right) + \Delta \langle L, M\rangle_{T_k} \left(\sum_{l=1}^{i\wedge k} \left(1 - \frac{l}{i}\right) \Delta L_{T_{k-l}} \right) \right) \right)^2 \right]$$
$$\le \frac{N}{iT} C \delta_N^2 \mathbb{E}\left[\left(\sum_{T_k \le t} \left(\sum_{l=1}^{i\wedge k} \left(1 - \frac{l}{i}\right) \Delta L_{T_{k-l}} + \sum_{l=1}^{i\wedge k} \left(1 - \frac{l}{i}\right) \Delta M_{T_{k-l}} \right) \right)^2 \right] = \mathcal{O}\left(iN\delta_N^2\right) = \scriptstyle{\mathcal{O}}(1).$$

The order follows from the evaluation of the second moment that has been carried out for the calculation of $\langle \mathfrak{D}^N\rangle_t$ before. For this reason the quadratic covariation $\langle \mathfrak{D}^N, L\rangle$ converges to zero in probability on $[0,T]$. It can be directly deduced that $\langle \mathfrak{D}^N, M\rangle_t = \scriptstyle{\mathcal{O}}_p(1)$ as well. If $L^\perp$ is a bounded $(\mathcal{F}_t)$-martingale with $\langle L, L^\perp\rangle \equiv 0$, the quadratic covariation

$$\langle \mathfrak{D}^N, L^\perp\rangle_t = \sqrt{\frac{N}{iT}} \sum_{T_k \le t} \Delta \langle L^\perp, M\rangle_{T_k} (\sum_{l=1}^{i\wedge k} \left(1 - \frac{l}{i}\right) \Delta L_{T_{k-l}})$$

converges to zero in probability on $[0,T]$ what is concluded analogously. For every bounded $(\mathcal{F}_t)$-martingale $M^\perp$ with $\langle M, M^\perp\rangle \equiv 0$ it also holds true that $\langle \mathfrak{D}^N, M^\perp\rangle_t \xrightarrow{p} 0$ for all $t \in [0,T]$. An application of Jacod's Theorem 1.6 in the same manner as for the proof of Theorem 3.1 completes the proof of Proposition 4.3.7. □
The stable central limit theorem for the one-scale subsampling estimator in Proposition 4.3.5 is implied from the marginal distribution for $t = T$.

Proposition 4.3.11. *On the same assumptions as in Proposition 4.3.6, the continuous martingale*

$$\mathfrak{M}_t^N := \sqrt{\frac{N}{M_N}} \sum_{i=1}^{M_N} \left[\sum_{T_k \le t} \left(\Delta L_{T_k} + \int_{T_k}^t \sigma_s^X dW_s^X \right) \left(\sum_{l=1}^{i\wedge k} \left(1 - \frac{l}{i}\right) \Delta M_{T_{k-l}} \right) \right.$$
$$\left. + \sum_{T_k \le t} \left(\Delta M_{T_k} + \int_{T_k}^t \sigma_s^Y dW_s^Y \right) \left(\sum_{l=1}^{i\wedge k} \left(1 - \frac{l}{i}\right) \Delta L_{T_{k-l}} \right) \right]$$

for $t \in [0,T]$ converges stably in law as $N \to \infty, M_N \to \infty, M_N/N^\alpha \to 0$ for every

$\alpha > 2/3$ to a limiting time-changed Brownian motion

$$\mathfrak{M}_t^N \overset{st}{\rightsquigarrow} \int_0^t \sqrt{v_{\mathfrak{M}_s}} d\tilde{\mathfrak{W}}_s^{\perp} ,$$

where $\tilde{\mathfrak{W}}^{\perp}$ is independent of $\mathcal{F}$ and with

$$v_{\mathfrak{M}_s} = \frac{26}{35} T G'(s)(\sigma_s^X \sigma_s^Y)^2 (1+\rho_s^2) .$$

Proof of Proposition 4.3.11:
The discretization error of the generalized multiscale estimator calculated with the closest synchronous approximation under the equivalent martingale measure where the drift terms equal zero

$$\sum_{i=1}^{M_N} \frac{\alpha_{i,M_N}^{opt}}{i} \sum_{j=i}^{N} \left(L_{T_j} - L_{T_{j-i}}\right)\left(M_{T_j} - M_{T_{j-i}}\right) - \langle X, Y \rangle_T$$
$$= \sum_{i=1}^{M_N} \alpha_{i,M_N}^{opt} \left(\frac{1}{i} \sum_{j=i}^{N} \left(L_{T_j} - L_{T_{j-i}}\right)\left(M_{T_j} - M_{T_{j-i}}\right) - \langle X, Y \rangle_T \right)$$

equals the weighted sum of $M_N \to \infty$ discretization errors of the type considered in Proposition 4.3.5 because $\sum_{i=1}^{M_N} \alpha_{i,M_N}^{opt} = 1$. Note, that all approximation errors in the preceding proof of Proposition 4.3.7 converge to zero in probability as long as $N \to \infty, i/N^{\alpha} \to 0$ for every $\alpha > 2/3$.
As outlined in the last section we begin with the proof of a multivariate stable central limit theorem for a finite-dimensional vector.

Lemma 4.3.12. *Consider the sequence of K-dimensional vectors $\mathbb{D}^N = \left(\mathfrak{D}^{i_N^1}, \ldots, \mathfrak{D}^{i_N^K}\right)$ where the entries $\mathfrak{D}^{i_N^k}, k = 1, \ldots, K < \infty$ are the continuous martingales*

$$\mathfrak{D}_t^{i_N^k} = \left[\sum_{T_r \le t} \left(\Delta L_{T_r} + \int_{T_r}^t \sigma_s^X dW_s^X\right) \left(\sum_{l=1}^{i_N^k \wedge r} \left(1 - \frac{l}{i_N^k}\right) \Delta M_{T_{r-l}} \right) \right.$$
$$\left. + \sum_{T_r \le t} \left(\Delta M_{T_r} + \int_{T_r}^t \sigma_s^Y dW_s^Y\right) \left(\sum_{l=1}^{i_N^k \wedge r} \left(1 - \frac{l}{i_N^k}\right) \Delta L_{T_{r-l}} \right) \right]$$

with a sequence of integers $i_N^k, k = 1, \ldots, K$. On the Assumptions 1, 4.1 and 4.3 and if for every $k \in \{1, \ldots, K\}$ there exists a constant q_k with $i_N^k / M_N \to q_k$, the following stable convergence holds true as $N \to \infty, M_N \to \infty, M_N/N^{\alpha} \to 0$ for every $\alpha > 2/3$:

$$\sqrt{\frac{N}{M_N}} \mathbb{D}_t^N \overset{st}{\rightsquigarrow} \int_0^t w_s d\mathbb{W}_s , \tag{4.21}$$

with a K-dimensional Brownian motion $\mathbb{W}$ independent of $\mathcal{F}$ and a predictable process

w_s with

$$(w_s w_s^*)_{mn} = \frac{T}{3} \min(q_m, q_n) \left(3 - \frac{\min(q_m, q_n)}{\max(q_m, q_n)}\right) (1 + \rho_s^2) \left(\sigma_s^X \sigma_s^Y\right)^2 G'(s) \tag{4.22}$$

with the convention that for $q_m = q_n = 0$ the ratio is one.
For $\mathbb{D}_T^N$ we obtain the following multivariate stable central limit theorem

$$\sqrt{\frac{N}{M_N}} \mathbb{D}_T^N \overset{st}{\rightsquigarrow} \mathbf{N}\left(0, \eta^2 \Sigma\right) \ , \tag{4.23}$$

with $\eta^2 = 2T \int_0^T (1 + \rho_t^2)(\sigma_t^X \sigma_t^Y)^2 G'(t) dt$ and

$$\Sigma_{mn} = \frac{1}{6} \min(q_m, q_n) \left(3 - \frac{\min(q_m, q_n)}{\max(q_m, q_n)}\right) \ .$$

Proof. Define for $k \in \{1, \ldots, K\}$ the continuous martingales

$$\mathfrak{M}_t^{i_N^k} = \sqrt{\frac{N}{M_N}} \mathfrak{D}_t^{i_N^k} \ .$$

For one single fixed k this type of martingales corresponding to the discretization error of the closest synchronous approximation of a one-scale subsampling estimator with subsampling frequency i_N^k has been considered in Proposition 4.3.7. We already know that

$$\langle \mathfrak{M}^{i_N^k} \rangle_t \xrightarrow{p} \frac{2}{3} T q_k \int_0^t (1 + \rho_s^2)(\sigma_s^X \sigma_s^Y)^2 G'(s) \, ds$$

from the proof of this Proposition where q_k additionally appears in the stochastic limit since $M_N T$ is the denominator in the root of the factor instead of $i_N^k T$.
The limit of the quadratic covariations $\langle \mathfrak{M}^{i_N^m}, \mathfrak{M}^{i_N^k} \rangle$ is derived using the same approximations as for the quadratic variation in the proof of Proposition 4.3.7. It follows that

$$\begin{aligned}
\langle \mathfrak{M}^{i_N^m}, \mathfrak{M}^{i_N^k} \rangle_t = \frac{N}{M_N} \Bigg[& \sum_{T_r \le t} \Delta \langle L \rangle_{T_r} \left(\sum_{l=1}^{\min(i_N^m, i_N^k, r)} \left(1 - \frac{l}{i_N^m}\right) \left(1 - \frac{l}{i_N^k}\right) \left(\Delta M_{T_{r-l}}\right)^2 \right) \\
& + \sum_{T_r \le t} 2 \Delta \langle L, M \rangle_{T_r} \left(\sum_{l=1}^{\min(i_N^m, i_N^k, r)} \left(1 - \frac{l}{i_N^m}\right) \left(1 - \frac{l}{i_N^k}\right) \Delta L_{T_{r-l}} \Delta M_{T_{r-l}} \right) \\
& + \sum_{T_r \le t} \Delta \langle M \rangle_{T_r} \left(\sum_{l=1}^{\min(i_N^m, i_N^k, r)} \left(1 - \frac{l}{i_N^m}\right) \left(1 - \frac{l}{i_N^k}\right) \left(\Delta L_{T_{r-l}}\right)^2 \right) + \mathcal{O}_p(1) \Bigg] \\
= N & \sum_{T_r \le t} 2 \, \frac{G^{(N)}(T_r) - G^{(N)}(T_{r-1})}{\Delta T_r} (\rho_{T_{r-1}}^2 + 1)(\sigma_{T_{r-1}}^X \sigma_{T_{r-1}}^Y)^2 \Delta T_r
\end{aligned}$$

$$\times \left(\sum_{l=1}^{\min(i_N^m, i_N^k, r)} \left(1 - \frac{l}{i_N^m}\right) \left(1 - \frac{l}{i_N^k}\right) \right) + \mathcal{O}_p(1)$$

$$\xrightarrow{p} 2T \int_0^t (\rho_s^2 + 1)(\sigma_s^X \sigma_s^Y)^2 G'(s) \left(\frac{1}{6} \min(q_m, q_k) \left(3 - \frac{\min(q_m, q_k)}{\max(q_m, q_k)} \right) \right) ds$$

since $\sum_{l=1}^{m}(1-(l/m))(1-(l/M)) = (1/2)m - (m^2/6M) - 1/8 + 1/(12M)$ for $m, M \in \mathbb{Z}$. The convergence in probability of the quadratic (co-)variations and a tightness criterion on $\langle \mathbb{D}^N, \mathbb{D}^{N\,*} \rangle$, which is fulfilled here, are enough to establish weak convergence of the vector (cf. Corollary VI. 6. 30 in Jacod and Shiryaev [2003]).
Corollary 1.2.4 enables us to prove the stronger result of stable weak convergence of the vector provided we can verify conditions (1.6a) and (1.6b). Since the reference continuous martingales for all entries of the vector $\mathbb{D}^N$ are L and M, it is sufficient to show that

$$\langle \mathbb{D}^N, \mathbb{L} \rangle_t \xrightarrow{p} 0 \quad , \quad \langle \mathbb{D}^N, \mathbb{M} \rangle_t \xrightarrow{p} 0, \quad \forall t \in [0, T] \; ,$$

where $\mathbb{L}$ denotes the vector with entries $\mathbb{L}^j = L, j = 1, \ldots, K$ and $\mathbb{M}$ with $\mathbb{M}^j = M, j = 1, \ldots, K$, respectively, and

$$\langle \mathbb{D}^N, \mathbb{L}^{\perp} \rangle_t \xrightarrow{p} 0 \quad , \quad \langle \mathbb{D}^N, \mathbb{M}^{\perp} \rangle_t \xrightarrow{p} 0, \quad \forall t \in [0, T]$$

where $\mathbb{L}^{\perp}$ and $\mathbb{M}^{\perp}$ are bounded $(\mathcal{F}_t)$-adapted martingales orthogonal to $\mathbb{L}$ and $\mathbb{M}$, respectively. The componentwise proof of the conditions above is the same as for the univariate case in the proof of Proposition 4.3.7. We conclude that conditions (1.5), (1.6a) and (1.6b) are fulfilled and Corollary 1.2.4 yields the result of the lemma. The asymptotic distribution of the vector is described by a limiting Brownian motion on $[0, T]$, and the marginal distribution at time T by a mixed Gaussian limit, where the normal distribution is defined as well as for all componentwise marginals on an orthogonal extension of the original underlying probability space. □

From the preceding multivariate limit theorem the Cramér-Wold device allows to conclude the weak convergence of all one-dimensional linear combinations of the transformed discretization errors of a finite collection of one-scale subsampling estimators. In particular, a weighted sum of the type of our generalized multiscale approach weakly converges. For an asymptotically $\mathbf{N}(0, \Sigma)$-distributed random vector the sum of all components is asymptotically normally distributed with variance $\sum_{i,j}(\Sigma_{ij})$ by the Cramér-Wold device and the normality of any linear sum of components of a multivariate normal distribution (see e. g. pp.516-517 in Rao [2001]).
The asymptotic variance in Proposition 4.3.6 is deduced from the multivariate limit and

$$\sum_{k,l} (\Sigma_{k,l}) = 2 \sum_{k=1}^{M_N} \sum_{l=1}^{k} \frac{l}{6M_N} \left(3 - \frac{l}{k}\right) \alpha_{k,M_N}^{opt} \alpha_{l,M_N}^{opt} + \mathcal{O}(1) = \frac{13}{35} + \mathcal{O}(1)$$

with the weights (4.14).
For the completion of the proof of Propositions 4.3.11 and 4.3.6, it remains to extend the

result for asymptotically infinitely many addends. This part of the proof uses similar methods as in Zhang [2006] where a central limit theorem for a multiscale estimator for the integrated volatility in the univariate setting is proved.
Let $0 < \delta < 1$ be an arbitrarily chosen real number and $\alpha = 1 - \delta/\sqrt{2}$. The ansatz is to approximate an infinite sum by finite sums incorporating $J := \min\{n \in \mathbb{N} | 2\alpha^{n-1} \leq \delta^2\}$ addends. For a subsampling frequency i_N^k choose $\tilde{\imath}_N^k \in \{1, \ldots, J\}$ such that $i_N^k/M_N \in (\alpha^{J-\tilde{\imath}_N^k - 1}, \alpha^{J-\tilde{\imath}_N^k})$. For our generalized multiscale estimator the subsampling frequencies follow the regular scheme $i_N^k = k \in \{1, \ldots, M_N\}$.
Now, let $i_N^* \in \{1, \ldots, M_N\}$ be the subsampling frequency for which the variance of the approximation errors for the one-scale estimators is maximized for fixed t:

$$i_N^* := \arg\max_{i \in \{1,\ldots,M_N\}} \mathbb{E}\left[\left(\widehat{\langle X, Y\rangle}_t^{sub,i_N} - \widehat{\langle X, Y\rangle}_t^{sub,\tilde{\imath}_N}\right)^2\right],$$

For an unbounded set of integers $\mathcal{N}$ for which $(i_n/M_n, i_n^*/M_n)_{n\in\mathcal{N}}$ converges to a limit (κ_1, κ_2), the multivariate limit theorem above yields

$$\begin{aligned}
&\frac{n}{M_n}\mathbb{E}\left[\left(\widehat{\langle X, Y\rangle}_t^{sub,i_n} - \widehat{\langle X, Y\rangle}_t^{sub,i_n^*}\right)^2\right] \\
&\qquad \longrightarrow \mathbb{E}\eta^2\left(\frac{\kappa_1}{3} + \frac{\kappa_2}{3} - 2\frac{1}{6}\kappa_1\left(3 - \frac{\kappa_1}{\kappa_2}\right)\right) \\
&\qquad = \mathbb{E}\eta^2\frac{1}{3}\kappa_2\left(1 - \frac{\kappa_1}{\kappa_2}\right)^2 \leq \frac{1}{3}\mathbb{E}\eta^2\delta^2 \ .
\end{aligned}$$

Since every subsequence has a subsequence for which $(i_n/M_n, i_n^*/M_n)$ converges, the result follows from

$$\begin{aligned}
&\frac{N}{M_N}\mathbb{E}\left[\left(\sum_{i=1}^{M_N}\left(\alpha_{i,M_N}^{opt}\right)^2\left(\left(\widehat{\langle X, Y\rangle}_t^{sub,i_N} - \widehat{\langle X, Y\rangle}_t^{sub,\tilde{\imath}_N}\right)\right)^2\right)^2\right] \\
&\qquad \leq C\delta^2\left(\sum_{i=1}^{M_N}\left(\alpha_{i,M_N}^{opt}\right)^2 + \sum_{i,k}\alpha_{i,M_N}^{opt}\alpha_{k,M_N}^{opt}\right) \\
&\qquad = \mathcal{O}\left(\delta^2\right)
\end{aligned}$$

and then letting $\delta \to 0$. This completes the proof of Propositions 4.3.11 and 4.3.6. The stability conditions (1.3a) and (1.3b) for the sum follow directly from the corresponding conditions that have been proved to show 4.3.12. □

Discretization error due to the lack of synchronicity

We take up the illustration (4.12) of the leading term of the error due to interpolations at the times $T_i^{(N)}, i = 0, \ldots, N$ in the discretization part of the estimation error of a one-scale subsampling estimator. We start with analyzing the error for the one-scale estimator (4.3) and then consider the multiscale extension (4.2) from Section 4.2.
Denote L and M the transformed efficient processes under the equivalent martingale measure where drifts are zero as in the foregoing proof. Define the corresponding transformed term

$$\begin{aligned}\mathcal{A}_T^N = \frac{1}{i}\sum_{j=i}^{N-1}\Bigg[&\left(L_{g_j} - L_{T_j}\right)\left(\sum_{k=j-i+1}^{j}\Delta M_{T_k}\right) + \left(M_{\gamma_j} - M_{T_j}\right)\left(\sum_{k=j-i+1}^{j}\Delta L_{T_k}\right)\\ &+\Delta L_{T_{j+1}}\left(\sum_{k=j-i+1}^{j}\left(M_{T_k} - M_{\lambda_{k+1}}\right)\right) + \Delta M_{T_{j+1}}\left(\sum_{k=j-i+1}^{j}\left(L_{T_k} - L_{l_{k+1}}\right)\right)\Bigg]\,.\end{aligned}$$

The outer sum above is centred and a sum of uncorrelated random variables since previous- and next-tick interpolations of one process do not overlap and, furthermore, $T_i^{(N)} \leq g_i^{(N)} \leq T_{i+1}^{(N)}, i = 0, \ldots, N$ holds. This will be thoroughly proved in the successive Proposition 4.3.14. First, we reveal that the addends including drift terms in (4.12) are of smaller order than the addends only including Brownian increments in the above term.

Lemma 4.3.13. *On the Assumptions 1 and 4.1, it holds true that*

$$\begin{aligned}&\frac{1}{i}\sum_{j=i}^{N-1}\Bigg[\left(X_{g_j} - X_{T_j}\right)\left(\sum_{k=j-i+1}^{j}\Delta Y_{T_k}\right) - \int_{T_j}^{g_j}\sigma_t^X\,dB_t^X\int_{T_{j-i}}^{T_j}\sigma_t^Y\,dB_t^Y\\ &\quad+\left(Y_{\gamma_j} - Y_{T_j}\right)\left(\sum_{k=j-i+1}^{j}\Delta X_{T_k}\right) - \int_{T_j}^{g_j}\sigma_t^Y\,dB_t^Y\int_{T_{j-i}}^{T_j}\sigma_t^X\,dB_t^X\\ &\quad+\Delta X_{T_{j+1}}\left(\sum_{k=j-i+1}^{j}\left(Y_{T_k} - Y_{\lambda_{k+1}}\right)\right) - \int_{T_j}^{T_{j+1}}\sigma_t^X\,dB_t^X\sum_{k=j-i+1}^{j}\int_{\lambda_{k+1}}^{T_k}\sigma_t^Y\,dB_t^Y\\ &\quad+\Delta Y_{T_{j+1}}\left(\sum_{k=j-i+1}^{j}\left(X_{T_k} - X_{l_{k+1}}\right)\right) - \int_{T_j}^{T_{j+1}}\sigma_t^Y\,dB_t^Y\sum_{k=j-i+1}^{j}\int_{\lambda_{k+1}}^{T_k}\sigma_t^X\,dB_t^X\Bigg]\\ &= \mathcal{O}_p(N^{-1})\,.\end{aligned}$$

Proof. Terms of the type $\int_{T_j}^{g_j}\mu_t^X\,dt\int_{T_{j-i}}^{T_j}\mu_t^Y\,dt\,, i \leq j \leq N$ have an expectation bounded by a constant multiplied with the time instant and second moment at most of order $i^2\delta_N^4$. Therefore, the whole term above has an expectation of order N^{-1} since the mixed addends with Brownian increments are centred. Terms of the type $\int_{T_j}^{g_j}\sigma_t^X\,dB_t^X\int_{T_{j-i}}^{T_j}\mu_t^Y\,dt$ have second moments at most of order $i^2\delta_N^3$ and terms of the type $\int_{T_j}^{g_j}\mu_t^X\,dt\int_{T_{j-i}}^{T_j}\sigma_t^Y\,dB_t^Y$ of order $i\delta_N^3$. The variance of the whole term is hence at most of order $i^{-2}N^2\cdot i^2\delta_N^3 = N^2\delta_N^3$.

These findings imply the statement of the lemma. □

Proposition 4.3.14. *On the Assumptions 1 and 4.1 the term $\mathcal{A}_T^N$ is the endpoint of a discrete martingale with respect to the filtration $\mathcal{F}_{j,N} = \mathcal{F}_{T_{j+1}^{(N)}}$. It holds true that*

$$\mathcal{A}_T^N = \mathcal{O}_p\left(i^{-1/2}N^{-1/2}\right) \ .$$

Proof.

$$\frac{1}{i}\Bigg[\left(L_{g_j}-L_{T_j}\right)\left(\sum_{k=j-i+1}^{j}\Delta M_{T_k}\right)+\left(M_{\gamma_j}-M_{T_j}\right)\left(\sum_{k=j-i+1}^{j}\Delta L_{T_k}\right)$$
$$+\Delta L_{T_{j+1}}\left(\sum_{k=j-i+1}^{j}\left(M_{T_k}-M_{\lambda_{k+1}}\right)\right)+\Delta M_{T_{j+1}}\left(\sum_{k=j-i+1}^{j}\left(L_{T_k}-L_{l_{k+1}}\right)\right)\Bigg]$$

for $i \le j \le N-1$ is $\mathcal{F}_{j,N} = \mathcal{F}_{T_{j+1}^{(N)}}$-measurable and the conditional expectation equals zero:

$$\frac{1}{i}\mathbb{E}\Bigg[\left(L_{g_j}-L_{T_j}\right)\left(\sum_{k=j-i+1}^{j}\Delta M_{T_k}\right)+\left(M_{\gamma_j}-M_{T_j}\right)\left(\sum_{k=j-i+1}^{j}\Delta L_{T_k}\right)$$
$$+\Delta L_{T_{j+1}}\left(\sum_{k=j-i+1}^{j}\left(M_{T_k}-M_{\lambda_{k+1}}\right)\right)+\Delta M_{T_{j+1}}\left(\sum_{k=j-i+1}^{j}\left(L_{T_k}-L_{l_{k+1}}\right)\right)|\mathcal{F}_{T_j^{(N)}}\Bigg]$$
$$=\frac{1}{i}\Bigg(\mathbb{E}\left[L_{g_j}-L_{T_j}\right]\left(\sum_{k=j-i+1}^{j}\Delta M_{T_k}\right)+\mathbb{E}\left[M_{\gamma_j}-M_{T_j}\right]\left(\sum_{k=j-i+1}^{j}\Delta L_{T_k}\right)$$
$$+\mathbb{E}\left[\Delta L_{T_{j+1}}\right]\left(\sum_{k=j-i+1}^{j}\left(M_{T_k}-M_{\lambda_{k+1}}\right)\right)+\mathbb{E}\left[\Delta M_{T_{j+1}}\right]\left(\sum_{k=j-i+1}^{j}\left(L_{T_k}-L_{l_{k+1}}\right)\right)\Bigg)=0.$$

The martingale is centred, since the addends incorporate products of Brownian increments over disjoint time intervals. The conditional variance yields

$$\frac{1}{i^2}\sum_{j=i}^{N-1}\mathbb{E}\Bigg[\Bigg(\left(L_{g_j}-L_{T_j}\right)\left(\sum_{k=j-i+1}^{j}\Delta M_{T_k}\right)+\left(M_{\gamma_j}-M_{T_j}\right)\left(\sum_{k=j-i+1}^{j}\Delta L_{T_k}\right)$$
$$+\Delta L_{T_{j+1}}\left(\sum_{k=j-i+1}^{j}\left(M_{T_k}-M_{\lambda_{k+1}}\right)\right)+\Delta M_{T_{j+1}}\left(\sum_{k=j-i+1}^{j}\left(L_{T_k}-L_{l_{k+1}}\right)\right)\Bigg)^2|\mathcal{F}_{T_j^{(N)}}\Bigg]$$
$$=\frac{1}{i^2}\sum_{j=i}^{N-1}\Bigg(\mathbb{E}\left[(L_{g_j}-L_{T_j})^2\right]\left(\sum_{k=j-i+1}^{j}\Delta M_{T_k}\right)^2+\mathbb{E}\left[(M_{\gamma_j}-M_{T_j})^2\right]\left(\sum_{k=j-i+1}^{j}\Delta L_{T_k}\right)^2$$

$$
\begin{aligned}
&+\mathbb{E}\left[(\Delta L_{T_{j+1}})^2\right]\left(\sum_{k=j-i+1}^{j}\left(M_{T_k}-M_{\lambda_{k+1}}\right)\right)^2\\
&+\mathbb{E}\left[(\Delta M_{T_{j+1}})^2\right]\left(\sum_{k=j-i+1}^{j}\left(L_{T_k}-L_{l_{k+1}}\right)\right)^2\\
&+\mathbb{E}\left[\int_{T_j}^{g_j}(\sigma_t^X)^2dt\right]\left(\sum_{k=j-i+1}^{j}\left(M_{T_k}-M_{\lambda_{k+1}}\right)\right)\left(\sum_{k=j-i+1}^{j}\Delta M_{T_k}\right)\\
&+\mathbb{E}\left[\int_{T_j}^{\gamma_j}(\sigma_t^Y)^2dt\right]\left(\sum_{k=j-i+1}^{j}\left(L_{T_k}-L_{l_{k+1}}\right)\right)\left(\sum_{k=j-i+1}^{j}\Delta L_{T_k}\right)\\
&+\mathbb{E}\left[\int_{T_j}^{g_j}\rho_t\sigma_t^X\sigma_t^Y\,dt\right]\left(\sum_{k=j-i+1}^{j}\left(L_{T_k}-L_{l_{k+1}}\right)\right)\left(\sum_{k=j-i+1}^{j}\Delta M_{T_k}\right)\\
&+\mathbb{E}\left[\int_{T_j}^{\gamma_j}\rho_t\sigma_t^X\sigma_t^Y\,dt\right]\left(\sum_{k=j-i+1}^{j}\left(M_{T_k}-M_{\lambda_{k+1}}\right)\right)\left(\sum_{k=j-i+1}^{j}\Delta L_{T_k}\right)\Bigg)\\
&=\mathcal{O}_p\left(i^{-1}N^{-1}\right)\ .
\end{aligned}
$$

The variance of the term is of order $(iN)^{-1}$ which can be proved by taking the expectation of the above given conditional variance and an upper bound of the second moment. We have used Itô isometry in the above calculation. The asymptotic orders of the addends follow from taking the expectations using Itô isometry and analyzing the differences of the addends minus their expectations, that converge to zero at a faster rate. This is analogous to the proofs for the discretization error of the closest synchronous approximation and we forgo a more detailed computation here. □

Denote $A_T^{N,i}$ the error due to non-synchronicity and interpolations for a fixed subsampling frequency $i=1,\ldots,M_N$ in the following. The error due to asynchronicity of the generalized multiscale estimator (4.2) equals the weighted sum $\sum_{i=1}^{M_N}\alpha_{i,M_N}^{opt}A_T^{N,i}$. We obtain directly that drift terms are asymptotically negligible again. The term $\sum_{i=1}^{M_N}\alpha_{i,M_N}^{opt}\mathcal{A}_T^{N,i}$ has expectation zero and the variance is of order

$$
\begin{aligned}
\mathbb{V}\mathrm{ar}\left(\sum_{i=1}^{M_N}\alpha_{i,M_N}^{opt}\mathcal{A}_T^{N,i}\right)&=\sum_{i,k}\alpha_{i,M_N}^{opt}\alpha_{k,M_N}^{opt}\mathbb{C}\mathrm{ov}\left(\mathcal{A}_T^{N,i}\,,\,\mathcal{A}_T^{N,k}\right)\\
=\underbrace{\sum_{i=1}^{M_N}\left(\alpha_{i,M_N}^{opt}\right)^2\mathbb{E}\left[\left(\mathcal{A}_T^{N,i}\right)^2\right]}_{=\mathcal{O}\left(M_N^{-2}N^{-1}\right)}&+\underbrace{\sum_{i\neq k}\alpha_{i,M_N}^{opt}\alpha_{k,M_N}^{opt}\mathbb{E}\left[\mathcal{A}_T^{N,i}\mathcal{A}_T^{k,N}\right]}_{=\mathcal{O}\left(M_N^{-1}N^{-1}\right)}=\mathcal{O}\left(\frac{M_N}{N}\right)\ .
\end{aligned}
$$

Thus, the error due to interpolations is of smaller asymptotic order than the discretization error of the closest synchronous approximation and asymptotically negligible.

4.3.3 Asymptotics of the cross term

For a one-scale subsampling estimator cross terms are asymptotically negligible and hence the stable central limit theorem in Theorem 4.2.2 is implied by Theorem 4.3.3. For the proof of the stable central limit theorem in Theorem 4.1 for the multiscale approach, we cope with the asymptotics of the cross terms in this subsection.

Proposition 4.3.15. *On the Assumptions 1, 4.1, 4.2 and 3, the cross terms of the generalized multiscale estimator* (4.2) *with noise-optimal weights* (4.14) *weakly converge to a mixed normal limit as* $M_N \to \infty$, $N \to \infty$, $M_N \delta_N \to 0$*:*

$$\sqrt{M_N}\sum_{i=1}^{M_N}\frac{\alpha_{i,M_N}^{opt}}{i}\sum_{j=i}^{N}\Big((X_{g_j}-X_{l_{j-i+1}})(\epsilon^Y_{\gamma_j}-\epsilon^Y_{\lambda_{j-i+1}})+(Y_{\gamma_j}-Y_{\lambda_{j-i+1}})(\epsilon^X_{g_j}-\epsilon^X_{l_{j-i+1}})\Big)$$
$$\rightsquigarrow \mathbf{N}\left(0,\mathbf{AVAR}_{cross}\right)\,, \tag{4.24}$$

with asymptotic variance

$$\mathbf{AVAR}_{cross}=\frac{12}{5}\left(\eta_Y^2\int_0^T(1+I_Y'(t))(\sigma_t^X)^2\,dt+\eta_X^2\int_0^T(1+I_X'(t))(\sigma_t^Y)^2\,dt\right)\,. \tag{4.25}$$

The convergence holds conditionally given the paths of the efficient processes.

Proof. This proof affiliates to the discussion in the preceding section, where degrees of regularity of non-synchronous sampling schemes have been defined in Definition 4.2.1 that are assumed to converge to continuously differentiable functions.
On the Assumption 3 of independent observation noise of X and Y, the two different cross terms are uncorrelated and we prove a central limit theorem for the first one:

$$\sqrt{M_N}\sum_{i=1}^{M_N}\frac{\alpha_{i,M_N}^{opt}}{i}\sum_{j=i}^{N}(X_{g_j}-X_{l_{j-i+1}})(\epsilon^Y_{\gamma_j}-\epsilon^Y_{\lambda_{j-i+1}})\rightsquigarrow\mathbf{N}\left(0,\frac{12}{5}\eta_Y^2\int_0^T(1+I_Y'(t))(\sigma_t^X)^2\,dt\right).$$

The parallel result for the other term can be proved analogously.
For the purpose of a shorter notation we have left out superscripts of the observation times, and write α_i, $i=1,\ldots,M_N$ for the weights although we are interested in the specific weights (4.14). Denote $\delta_N=\sup_{i\in\{1,\ldots,N\}}\Delta T_i$ and $\gamma_{j,+}=\min\left(\tau_k\in\mathcal{T}^Y|\tau_k\in\mathcal{G}^{j+1}\right)$, $g_{j,+}=\min\left(t_k\in\mathcal{T}^X|t_k\in\mathcal{H}^{j+1}\right)$ and C a generic constant as before. From

$$\mathbb{E}\left[\left(\sqrt{M_N}\sum_{i=1}^{M_N}\frac{\alpha_i}{i}\sum_{j=i}^{N}(X_{g_j}-X_{T_j})(\epsilon^Y_{\gamma_j}-\epsilon^Y_{\lambda_{j-i+1}})+(X_{T_{j-i}}-X_{l_{j-i+1}})(\epsilon^Y_{\gamma_j}-\epsilon^Y_{\lambda_{j-i+1}})\right)^2\right]$$
$$\le M_N\sum_{i,k\in\{1,\ldots,M_N\}}\frac{\alpha_i\alpha_k}{ik}2\eta_Y^2\left(\sum_{j=i\vee k}^{N}\mathbb{E}(X_{g_j}-X_{T_j})^2+\sum_{j=0}^{N-(i\vee k)}\mathbb{E}(X_{T_j}-X_{l_{j+1}})^2\right)$$

$$\leq M_N\, C4\eta_Y^2 \sum_{i,k\in\{1,\dots,M_N\}} \frac{\alpha_i\alpha_k}{ik} = \mathcal{O}\left(M_N^{-1}\right) \;,$$

for the errors due to interpolations and

$$\mathbb{E}\left[\left(\sqrt{M_N}\sum_{i=1}^{M_N}\frac{\alpha_i}{i}\left(\sum_{k=N-i+1}^{N}\epsilon^Y_{\gamma_k}(X_{T_k}-X_{T_{k-i}})-\sum_{k=1}^{i}\epsilon^Y_{\lambda_k}(X_{T_{k+i}}-X_{T_k})\right)\right)^2\right]$$
$$= M_N \sum_{i,k\in\{1,\dots,M_N\}}\frac{\alpha_i\alpha_k}{ik}\eta_Y^2\left(\sum_{r=N-(i\wedge k)+1}^{N}\mathbb{E}(X_{T_r}-X_{T_{r-i}})^2+\sum_{r=1}^{i\wedge k}\mathbb{E}(X_{T_{r+i}}-X_{T_r})^2\right)$$
$$= \mathcal{O}\left(M_N\delta_N\right)$$

for boundary terms, we conclude that

$$\sqrt{M_N}\sum_{i=1}^{M_N}\frac{\alpha_i}{i}\sum_{j=i}^{N}(X_{g_j}-X_{l_{j-i+1}})(\epsilon^Y_{\gamma_j}-\epsilon^Y_{\lambda_{j-i+1}})$$
$$= \sqrt{M_N}\sum_{i=1}^{M_N}\frac{\alpha_i}{i}\left(\sum_{j=i}^{N-i}\epsilon^Y_{\gamma_j}(X_{T_j}-X_{T_{j-i}})-\epsilon^Y_{\lambda_{j+1}}(X_{T_{j+i}}-X_{T_j})\right)+\mathcal{o}_p(1)$$
$$= \sqrt{M_N}\sum_{j=2}^{N-2}\left(\epsilon^Y_{\gamma_j}\sum_{i=1}^{M_N^*(j)}\frac{\alpha_i}{i}(X_{T_j}-X_{T_{j-i}})-\epsilon^Y_{\lambda_{j+1}}\sum_{i=1}^{M_N^*(j)}(X_{T_{j+i}}-X_{T_j})\right)+\mathcal{o}_p(1)$$
$$= \sqrt{M_N}\left(\sum_{j\in\mathcal{Y}_1}\epsilon^Y_{\gamma_j}\sum_{i=1}^{M_N^*(j)}\frac{\alpha_i}{i}\zeta^1_{i,j}+\sum_{j\in\mathcal{Y}_2}\epsilon^Y_{\gamma_j}\sum_{i=1}^{M_N^*(j)}\frac{\alpha_i}{i}\zeta^2_{i,j}\right.$$
$$+\sum_{j\in\mathcal{Y}_3}\epsilon^Y_{\gamma_j}\sum_{i=1}^{M_N^*(j)}\frac{\alpha_i}{i}\zeta^3_{i,j}+\sum_{j\in\mathcal{Y}_4}\epsilon^Y_{\gamma_j}\sum_{i=1}^{M_N^*(j)}\frac{\alpha_i}{i}\zeta^{4a}_{i,j}$$
$$\left.-\sum_{j\in\mathcal{Y}_4}\epsilon^Y_{\gamma_{j,+}}\sum_{i=1}^{M_N^*(j)}\frac{\alpha_i}{i}\zeta^{4b}_{i,j}\right)+\mathcal{o}_p(1)\;.$$

Here, we aggregate the observation times $\gamma_j,\lambda_j,\;j=2,\dots,N-2$ in disjoint sets conforming to the four cases introduced in the discussion of the preceding section. Denote thereto

$$\mathcal{Y}_1=\{j\in\{2,\dots,N_2\}|\gamma_j\neq\gamma_{j-1}\,,\,\gamma_j\leq g_j\}\;,$$
$$\mathcal{Y}_2=\{j\in\{2,\dots,N_2\}|\gamma_j> g_j\,,\,\gamma_j\geq g_{j,+}\}\;,$$
$$\mathcal{Y}_3=\{j\in\{2,\dots,N_2\}|\gamma_j> g_j\,,\,\gamma_j< g_{j,+}\,,\,\gamma_{j,+}>g_{j,+}\}\;,$$
$$\mathcal{Y}_4=\{j\in\{2,\dots,N_2\}|\gamma_j> g_j\,,\,\gamma_j< g_{j,+}\,,\,\gamma_{j,+}\leq g_{j,+}\}\;,$$

and $M_N^*(j)=\min\left(j,N-j,M_N\right)$. The increments of X that are multiplied with each observation error differ according to the set $\mathcal{Y}_k,\;1\leq k\leq 4$ to which γ_j belongs. We use the notation

$$\begin{aligned}
\zeta^1_{i,j} &= (X_{T_j} - X_{T_{j-i}}) - (X_{T_{j+i}} - X_{T_j})\ ,\\
\zeta^2_{i,j} &= (X_{T_j} - X_{T_{j-i}}) + (X_{T_{j+1}} - X_{T_{j-i+1}}) - (X_{T_{j+i+1}} - X_{T_{j+1}})\ ,\\
\zeta^3_{i,j} &= (X_{T_j} - X_{T_{j-i}}) - (X_{T_{j+i+1}} - X_{T_{j+1}})\ ,\\
\zeta^{4a}_{i,j} &= (X_{T_j} - X_{T_{j-i}})\ ,\ \zeta^{4b}_{i,j} = (X_{T_{j+i+1}} - X_{T_{j+1}})\ .
\end{aligned}$$

The resulting aggregated leading term above of the cross term is the endpoint of a discrete martingale with respect to the filtration $\mathcal{F}_{j,N} := \sigma\left(\epsilon^Y_{\tau_k^{(N)}} | \tau_k^{(N)} < \gamma_{j+1}^{(N)}\ , X, Y\right)$. Since if $j \in \mathcal{Y}_4 \Rightarrow \gamma_{j,+} < \gamma_{j+1}$, the martingale property with respect to the filtration $\mathcal{F}_{j,N}$ is assured by Assumption 3.
An application of Corollary 1.3.4 shows the asymptotic normality of the cross term conditionally on the paths of the efficient processes. The conditional Lindeberg condition can be verified (using Chebyshev's inequality or directly verifying the conditional Lyapunov condition) in the same way as in previous proofs and we omit it here. The sum of conditional variances yields

$$\begin{aligned}
&\sum_{l\in\{1,2,3,4a\}} \left(\sum_{j\in\mathcal{Y}_l} \mathbb{E}\Big[\Big(\sqrt{M_N}\epsilon^Y_{\gamma_j} \sum_{i=1}^{M_N^*(j)} \frac{\alpha_i}{i}\zeta^l_{i,j}\Big)^2 \Big|\mathcal{F}_{j-1,N}\Big]\right.\\
&\left.+\sum_{j\in\mathcal{Y}_4} \mathbb{E}\Big[\Big(-\sqrt{M_N}\epsilon^Y_{\gamma_{j,+}} \sum_{i=1}^{M_N^*(j)} \frac{\alpha_i}{i}\zeta^{4b}_{i,j}\Big)^2 \Big|\mathcal{F}_{j-1,N}\Big]\right)\\
&= M_N\eta_Y^2 \left(\sum_{j\in\mathcal{Y}_1\cup\mathcal{Y}_3\cup\mathcal{Y}_4} \left(\sum_{i=1}^{M_N^*(j)} \frac{\alpha_i}{i}\zeta^1_{i,j}\right)^2 + \sum_{j\in\mathcal{Y}_2}\left(\sum_{i=1}^{M_N^*(j)} \frac{\alpha_i}{i}\zeta^2_{i,j}\right)^2\right) + \mathcal{O}_p(1)\\
&= M_N\eta_Y^2 \left(\sum_{j\in\mathcal{Y}_1\cup\mathcal{Y}_3\cup\mathcal{Y}_4} \sum_{i,k} \frac{\alpha_i\alpha_k}{ik}\left(\zeta^1_{i\wedge k,j}\right)^2 + \sum_{j\in\mathcal{Y}_2}\left(\sum_{i,k} \frac{\alpha_i\alpha_k}{ik}\left(\zeta^2_{i\wedge k,j}\right)^2\right)\right) + \mathcal{O}_p(1)\\
&= M_N\eta_Y^2 \left(\sum_{j\in\mathcal{Y}_1\cup\mathcal{Y}_3\cup\mathcal{Y}_4} \sum_{i,k\in\{1,\ldots,M_N^*(j)\}} \frac{\alpha_i\alpha_k}{ik}\left((X_{T_j} - X_{T_{j-(i\wedge k)}})^2 + (X_{T_{j+(i\wedge k)}} - X_{T_j})^2\right)\right.\\
&\qquad\left.+\sum_{j\in\mathcal{Y}_2}\sum_{i,k} \frac{\alpha_i\alpha_k}{ik}\left(4(X_{T_j} - X_{T_{j-(i\wedge k)}})^2 + (X_{T_{j+(i\wedge k)}} - X_{T_j})^2\right)\right) + \mathcal{O}_p(1)\\
&= M_N\eta_Y^2 \sum_{j=2}^{N-2} \sum_{i,k\in\{1,\ldots,M_N^*(j)\}} \frac{\alpha_i\alpha_k}{ik}(2 + \mathbb{1}_{\{j\in\mathcal{Y}_2\}})(X_{T_j} - X_{T_{j-(i\wedge k)}})^2 + \mathcal{O}_p(1)\\
&= M_N\eta_Y^2 \left(2 \sum_{i,k\in\{1,\ldots,M_N^*(j)\}} \frac{\alpha_i\alpha_k}{ik}(i\wedge k)\widehat{\langle X\rangle}_T^{sub,i\wedge k}\right.
\end{aligned}$$

$$+ \sum_{i,k\in\{1,\dots,M_N^*(j)\}} \frac{\alpha_i\alpha_k}{ik} \sum_{j=i\wedge k}^{N} \mathbb{1}_{\{j\in\mathcal{Y}_2\}}(X_{T_j} - X_{T_{j-(i\wedge k)}})^2\Bigg) + \mathcal{O}_p(1)$$
$$\xrightarrow{p} \frac{12}{5}\eta_Y^2\left(\langle X\rangle_T + \int_0^T I_Y'(t)(\sigma_t^X)^2\,dt\right) .$$

Since for the shifted increments

$$(X_{T_{j+i+1}} - X_{T_{j+1}}) = (X_{T_{j+i}} - X_{T_j}) + \mathcal{O}_p\left(N^{-1/2}\right)$$

holds, where the order is for time instants of average length N^{-1}, the variances of the sums over all $j \in \mathcal{Y}_1$ and $j \in \mathcal{Y}_3$ are asymptotically equal. The variance of both uncorrelated sums over maxima γ_j and minima $\gamma_{j,+}$ distributed according to the fourth case is also asymptotically equal to the variances of those two addends. Only the asymptotic variance of the sum over all $j \in \mathcal{Y}_2$ is bigger. For this reason the total asymptotic variance hinges on the asymptotic degree of regularity of the non-synchronous sampling scheme $(\mathcal{T}^X, \mathcal{T}^Y)$ defined in Definition 4.2.1.
In the calculation of the asymptotic variance we have used that

$$\zeta_{i,j}^1\zeta_{i,k}^1 = \left(\zeta_{i\wedge k,j}^1\right)^2 + \zeta_{i\wedge k,j}^1\left(\sum_{l=j-(i\vee k)+1}^{j-(i\wedge k)} \Delta X_{T_l} + \sum_{l=j+(i\wedge k)+1}^{j+(i\vee k)} \Delta X_{T_l}\right) ,$$

where the second remainder addend has an expectation equal to zero, and analogous formulae for $\zeta_{i,j}^2$, for all $1 \le i \le M_N\,, 1 \le k \le M_N\,,\ k \vee i \le j \le N - (i \vee k)$.
Furthermore, an application of the mean value theorem, Itô isometry and approximations in the same spirit as in the calculation of the asymptotic variance in the proof of the central limit theorem for the discretization errors of the estimators, lead to the Riemann sum in the calculation of the asymptotic variance above. The cross terms in $(\zeta_{i,j}^l)^2$, $l = 1, 2$ are asymptotically negligible. Since in $\mathcal{Y}_4$ repeating maxima $\gamma_i = \gamma_{i+1}$ are considered only once, it holds true that $|\mathcal{Y}_1| + |\mathcal{Y}_3| + |\mathcal{Y}_4| + 2|\mathcal{Y}_2| = N - 3 \pm 1$ (the last addend can appear due to boundary term effects). In the last step we have used that

$$M_N \sum_{i,k\in\{1,\dots,M_N\}} \frac{\alpha_{i,M_N}^{opt}\alpha_{k,M_N}^{opt}}{ik}(i\wedge k) = 6/5 + \mathcal{o}(1)$$

when inserting the weights (4.14).
From the analysis for the asymptotic discretization error of a one-scale subsampling estimator, we know that

$$\widehat{\langle X\rangle}_T^{sub,i\wedge k} = \frac{1}{i\wedge k}\sum_{j=i\wedge k}^{N}(X_{T_j} - X_{T_{j-(i\wedge k)}})^2 = \langle X\rangle_T + \mathcal{O}_p\left(\sqrt{\frac{(i\wedge k)}{N}}\right)$$

holds true. Similarly, it can be deduced that

$$
\begin{aligned}
&\frac{1}{i}\sum_{j=i}^{N}\mathbb{1}_{\{j\in\mathcal{Y}_2\}}(X_{T_j}-X_{T_{j-i}})^2\\
&=\frac{1}{i}\sum_{l=1}^{N}(\Delta X_{T_l})^2\left(\sum_{k=1}^{i}\mathbb{1}_{\{(k+l-1)\in\mathcal{Y}_2\}}\right)+\mathcal{O}_p\left(\sqrt{i/N}\right)\\
&=\sum_{l=1}^{N}\left(\sigma^X_{T_{l-1}}\right)^2\Delta T_l\frac{\left(\sum_{k=1}^{i}\mathbb{1}_{\{(k+l-1)\in\mathcal{Y}_2\}}\right)}{T_{l+i-1}-T_{l-1}}\frac{T_{l+i-1}-T_{l-1}}{i}+\mathcal{O}_p\left(\sqrt{i/N}\right)\\
&=\sum_{l=1}^{N}\left(\sigma^X_{T_{l-1}}\right)^2\Delta T_l\frac{\left(\sum_{k=1}^{i}\mathbb{1}_{\{(k+l-1)\in\mathcal{Y}_2\}}\right)}{T_{l+i-1}-T_{l-1}}\frac{T}{N}+\mathcal{O}_p\left(\sqrt{i/N}\right)\\
&=\sum_{l=1}^{N}\left(\sigma^X_{T_{l-1}}\right)^2\Delta T_l\frac{\left(\sum_{k=1}^{i}\mathbb{1}_{\{(k+l-1)\in\mathcal{Y}_2\}}\right)}{T_{l+i-1}-T_{l-1}}\frac{T}{N}+\mathcal{O}_p\left(\sqrt{i/N}\right)\\
&=\sum_{l=1}^{N}\left(\sigma^X_{T_{l-1}}\right)^2\Delta T_l\frac{\left(I_Y^N(T_{l+i-1})-I_Y^N(T_{l-1})\right)}{T_{l+i-1}-T_{l-1}}+\mathcal{O}_p\left(\sqrt{i/N}\right)\\
&=\sum_{l=1}^{N}\left(\sigma^X_{T_{l-1}}\right)^2\Delta T_l\left(\frac{\left(I_Y^N(T_{l+i-1})-I_Y^N(T_{l-1})\right)}{T_{l+i-1}-T_{l-1}}-I_Y'(T_{l-1})\right)\\
&\quad+\sum_{l=1}^{N}\left(\sigma^X_{T_{l-1}}\right)^2\Delta T_l\;I_Y'(T_{l-1})+\mathcal{O}_p\left(\sqrt{i/N}\right)\\
&=\int_0^T I_Y'(t)(\sigma^X_t)^2\,dt+\mathcal{O}_p\left(\sqrt{i/N}\right)\;,
\end{aligned}
$$

on Assumption 4.2. The uniform convergence of the difference quotients and $\langle X\rangle_T = \int_0^T(\sigma^X_T)^2\,dt<\infty$ guarantee that the approximation error in the first addend converges to zero and, hence, the whole term to the corresponding integral of the Riemann sum in the second addend. The first equality in the above calculation is obtained in the same way as for the discretization error of the one-scale subsampling estimator. Squared increments $(\Delta X_{T_l})^2$, $l=1,\dots,N$ can appear in the subsampled squared increments $(X_{T_{(l+i-1)\wedge N}}-X_{T_{l-1}})^2,\dots,(X_{T_l}-X_{T_{(l-i)\wedge 0}})^2$. For the discretization error this leads to $i(\Delta X_{T_l})^2$ for $i\le l\le N-i+1$, but here only some of these increments are non-zero. The term of order $\sqrt{i/N}$ is due to the cross terms.
Proposition 4.3.15 follows from Corollary 1.3.4. □

5 Enhancement for feasible statistical inference and discussion of model specifications

5.1 Histogram-based consistent estimation of the asymptotic variances

The asymptotic variances (4.8) and (4.10) of the generalized multiscale estimator (4.2) and the one-scale subsampling estimator (4.3), and also the Hayashi-Yoshida estimator (3.2), appearing in the stable central limit theorems in Theorem 4.1, Corollary 4.2.2 and Theorem 3.1 in Sections 4.2 and 3.2, are random and depend on unknown quantities. In this section, we aim at estimating these asymptotic variances consistently that will make our limit theorems feasible. We consider the latent semimartingale model with i. i. d. noise contamination of Chapter 4, but also give an estimation method for the asymptotic variance of the Hayashi-Yoshida estimator in the asynchronous non-noisy setting at the end of this section in Proposition 5.1.7. If we are able to find consistent estimators $\widehat{\mathbf{AVAR}}_{multi}$ and $\widehat{\mathbf{AVAR}}_{sub}$ for the asymptotic variances, it holds true that

$$N^{1/4}\frac{\left(\widehat{\langle X,Y\rangle}_T^{multi}-\langle X,Y\rangle_T\right)}{\widehat{\mathbf{AVAR}}_{multi}}\stackrel{st}{\rightsquigarrow}\mathbf{N}(0,1)\ ,$$
$$N^{1/6}\frac{\left(\widehat{\langle X,Y\rangle}_T^{sub}-\langle X,Y\rangle_T\right)}{\widehat{\mathbf{AVAR}}_{sub}}\stackrel{st}{\rightsquigarrow}\mathbf{N}\left(0,1\right)\ .$$

This is grounded on the theory of stable convergence that has been discussed in Section 1.2 and constitutes the key element to draw statistical inference based on the mixed normality results deduced in the last chapter. We impose the same regularity assumptions as in the last chapter.

It is a well-known result in the field of integrated volatility estimation in the microstructure noise setting, that the noise variance can be estimated with the realized variance (cf. Zhang et al. [2005]):

$$\sqrt{n}\left((2n)^{-1}\sum_{i=1}^{n}(\Delta X_{t_i})^2-\eta_X^2\right)\rightsquigarrow\mathbf{N}\left(0,\mathbb{E}\left[\left(\epsilon_{t_1}^X\right)^4\right]\right)\ . \tag{5.1}$$

Furthermore, the estimators for η_X^2 and η_Y^2 are asymptotically uncorrelated on Assumption 3 since the uncorrelated noise terms dominate the correlated Brownian parts. The constant depending on the non-synchronous observation schemes in the variance due to noise $I_X(T)+I_Y(T)$ can be estimated taking simply the empirical values $I_X^N(T)+I_Y^N(T)$ that converge as $N\to\infty$ on the Assumption 4.2. In fact, these empirical versions are exactly the values influencing the non-asymptotic variance. Therefore, consistent estimators for the discretization variances and the variance due to cross terms for the multiscale estimator are required and lead to consistent estimators $\widehat{\mathbf{AVAR}}_{multi}$ and $\widehat{\mathbf{AVAR}}_{sub}$.
For a consistent estimation of the integrals

$$\begin{aligned}
\int_0^T G'(t)(\sigma_t^X\sigma_t^Y\rho_t)^2\,dt &= \int_0^{G(T)}\left(\sigma^X_{G^{-1}(u)}\sigma^Y_{G^{-1}(u)}\rho_{G^{-1}(u)}\right)^2\,du\ ,\\
\int_0^T G'(t)(\sigma_t^X\sigma_t^Y)^2\,dt &= \int_0^{G(T)}\left(\sigma^X_{G^{-1}(u)}\sigma^Y_{G^{-1}(u)}\right)^2\,du\ ,\\
\int_0^T I_Y'(t)(\sigma_t^X)^2\,dt &= \int_0^{I_Y(T)}\left(\sigma^X_{I_Y^{-1}(u)}\right)^2\,du\ ,\\
\int_0^T I_X'(t)(\sigma_t^Y)^2\,dt &= \int_0^{I_X(T)}\left(\sigma^Y_{I_X^{-1}(u)}\right)^2\,du\ ,
\end{aligned}$$

that appear in the asymptotic variance of the generalized multiscale estimator (4.8), we propose histogram-type estimators using bins according to timescales associated with the quadratic variation of synchronized sampling times and associated with the degree of regularity of asynchronicity, respectively. These sequences of functions have been defined in Section 4.2.
For this purpose, given a chosen number of bins K_N, with $K_N\to\infty$ and $K_N^{-1}N\to\infty$ as $N\to\infty$, we determine the assigned non-equispaced bin-widths $\Delta G_j^N = G_j^N - G_{j-1}^N$, $\Delta(I_X)_j^N = (I_X)_j^N - (I_X)_{j-1}^N$ and $\Delta(I_Y)_j^N = (I_Y)_j^N - (I_Y)_{j-1}^N$, $j\in\{1,\ldots,K_N\}$, where

$$\begin{aligned}
G_j^N &:= \inf\left\{t\in[0,T]\,\middle|\, G^N(t)=\frac{N}{T}\sum_{T_k^{(N)}\le t}\left(\Delta T_k^{(N)}\right)^2 \ge \frac{G^N(T)}{K_N}j\right\}\ ,\\
(I_X)_j^N &:= \inf\left\{t\in[0,T]\,\middle|\, I_X^N(t)=\frac{T}{N}\sum_{g_k^{(N)}\le t}\mathbb{1}_{\{g_k^{(N)}=g_{k-1}^{(N)}\}} \ge \frac{I_X^N(T)}{K_N}j\right\}\ ,\\
(I_Y)_j^N &:= \inf\left\{t\in[0,T]\,\middle|\, I_Y^N(t)=\frac{T}{N}\sum_{\gamma_k^{(N)}\le t}\mathbb{1}_{\{\gamma_k^{(N)}=\gamma_{k-1}^{(N)}\}} \ge \frac{I_Y^N(T)}{K_N}j\right\}\ ,
\end{aligned}$$

$j\in\{1,\ldots,K_N\}$, if $I_X^N(T)>0$ and $I_Y^N(T)>0$, respectively, $G_0^N=(I_X)_0^N=(I_Y)_0^N:=0$. Recall, that both, the quadratic variation of the sampling times of the closest synchronous approximation and the functions describing the degree of regularity of asynchronicity, are monotone increasing on $[0,T]$. For $I_X^N(T)=0$ or $I_Y^N(T)=0$, we estimate the

corresponding limiting functions to be zero.
On each bin we calculate estimators for the increase of the quadratic (co-) variations that are denoted $\widehat{\Delta\langle X\rangle}_{G_j^N}$, $\widehat{\Delta\langle Y\rangle}_{G_j^N}$, $\widehat{\Delta\langle X,Y\rangle}_{G_j^N}$, $\widehat{\Delta\langle X\rangle}_{(I_Y)_j^N}$ and $\widehat{\Delta\langle Y\rangle}_{(I_X)_j^N}$ in the following. Those estimators include noisy observations within the particular bin and are constructed as multiscale sums in the same spirit as the generalized multiscale estimator (4.2) for the whole time span $[0,T]$. As estimators for the quadratic variations we take analogous multiscale estimators, with the same weights (4.14), as introduced in Zhang [2006].
The underlain idea is to approximate the random functions $(\sigma_t^X\sigma_t^Y\rho_t)^2$, $(\sigma_t^X)^2$ and $(\sigma_t^Y)^2$, or rather the time transformed versions, by locally constant functions. The final time-adjusted histogram estimators

$$\hat{I}_1=\sum_{j=1}^{K_N}\left(\frac{\widehat{\Delta\langle X,Y\rangle}_{G_j^N}}{\Delta G_j^N}\right)^2\frac{G^N(T)}{K_N}\;, \tag{5.2a}$$

$$\hat{I}_2=\sum_{j=1}^{K_N}\left(\frac{\widehat{\Delta\langle X\rangle}_{G_j^N}\widehat{\Delta\langle Y\rangle}_{G_j^N}}{\left(\Delta G_j^N\right)^2}\right)\frac{G^N(T)}{K_N}\;, \tag{5.2b}$$

$$\hat{I}_3=\sum_{j=1}^{K_N}\left(\frac{\widehat{\Delta\langle X\rangle}_{(I_Y)_j^N}}{\Delta(I_Y)_j^N}\right)\frac{I_Y^N(T)}{K_N}\;, \tag{5.2c}$$

$$\hat{I}_4=\sum_{j=1}^{K_N}\left(\frac{\widehat{\Delta\langle Y\rangle}_{(I_X)_j^N}}{\Delta(I_X)_j^N}\right)\frac{I_X^N(T)}{K_N}\;, \tag{5.2d}$$

provide consistent estimators of the above integrals and can be used to obtain consistent estimators for the total asymptotic variances (4.8) and (4.10). The addends in (5.2a)-(5.2d) are uncorrelated since the observation errors are i. i. d. on Assumption 3 and Brownian increments over disjoint time intervals are independent and the drift terms are asymptotically negligible on Assumption 1.
At first, we examine the asymptotics of the addends in the above histogram-type sums and the corresponding multiscale estimators on bins. Throughout the next paragraph we use the notation $A_N\sim B_N$ to express that $A_N=\mathcal{O}(B_N)$ and $B_N=\mathcal{O}(A_N)$. Since $\lim_{N\to\infty}(G^N(T)/T)=G(T)/T>0$ holds, but $I_X(T)=0$ and $I_Y(T)=0$ is possible, the asymptotics for (5.2a) and (5.2b) are considered separately.

Corollary 5.1.1. *Let R_j^N denote the number of sampling times $T_k^{(N)}, 0\le k\le N$, of the closest synchronous approximation in one certain bin $[G_j^N,G_{j+1}^N), 0\le j\le K_N-1$, and R_j^n, R_j^m the number of observations of $\tilde{X}$ and $\tilde{Y}$ within the same time interval. Define the generalized multiscale estimator in the fashion of* (4.2)

$$\widehat{\Delta\langle X,Y\rangle}_{G_{j+1}^N}=\sum_{i=1}^{M_N(j)}\frac{\alpha_{i,M_N(j)}^{opt}}{i}\sum_{r=i}^{R_j^N}\left(\tilde{X}_{g_r}-\tilde{X}_{l_{r-i+1}}\right)\left(\tilde{Y}_{\gamma_r}-\tilde{Y}_{\lambda_{r-i+1}}\right)$$

for the increase of the quadratic covariation $\Delta\langle X,Y\rangle_{G^N_{j+1}}$ and the univariate multiscale estimators

$$\widehat{\Delta\langle X\rangle}_{G^N_{j+1}} = \sum_{i=1}^{M_n(j)} \frac{\alpha^{opt}_{i,M_n(j)}}{i} \sum_{r=i}^{R^n_j} \left(\tilde{X}_{t_r} - \tilde{X}_{t_{r-i}}\right)^2 \ ,$$
$$\widehat{\Delta\langle Y\rangle}_{G^N_{j+1}} = \sum_{i=1}^{M_m(j)} \frac{\alpha^{opt}_{i,M_m(j)}}{i} \sum_{r=i}^{R^m_j} \left(\tilde{Y}_{\tau_r} - \tilde{Y}_{\tau_{r-i}}\right)^2 \ .$$

For $M_N(j) = c_N(j)\cdot\sqrt{NK_N}, M_n(j) = c_n(j)\cdot\sqrt{NK_N}, M_m(j) = c_m(j)\cdot\sqrt{NK_N}$ with constants $c_N(j), c_n(j), c_m(j)$, the following limit theorems hold true:

$$N^{1/4}K_N^{1/4}\left(\widehat{\Delta\langle X,Y\rangle}_{G^N_{j+1}} - \Delta\langle X,Y\rangle_{G^N_{j+1}}\right) \overset{st}{\rightsquigarrow} \mathbf{N}\left(0,\eta_1^2\right) \ , \tag{5.3a}$$
$$N^{1/4}K_N^{1/4}\left(\widehat{\Delta\langle X\rangle}_{G^N_{j+1}} - \Delta\langle X\rangle_{G^N_{j+1}}\right) \overset{st}{\rightsquigarrow} \mathbf{N}\left(0,\eta_2^2\right) \ , \tag{5.3b}$$
$$N^{1/4}K_N^{1/4}\left(\widehat{\Delta\langle Y\rangle}_{G^N_{j+1}} - \Delta\langle Y\rangle_{G^N_{j+1}}\right) \overset{st}{\rightsquigarrow} \mathbf{N}\left(0,\eta_3^2\right) \ , \tag{5.3c}$$

with almost surely finite random asymptotic variances η_1^2, η_2^2 and η_3^2 as $N\to\infty$ with $K_N N^{-1/3}\to 0$.

Proof. The total estimation errors for each estimator can be split in four uncorrelated parts analogously as for $\widehat{\langle X,Y\rangle}_T^{multi}$ (cf. Section 4.2). Essential when considering the multiscale estimators on bins is that on Assumption 4.3 the distances between sampling times are of order $N^{-1}\sim n^{-1}\sim m^{-1}$, whereas the numbers of observations $R^N_j\sim R^n_j\sim R^m_j$ in the specific bin are of order NK_N^{-1}. Following the analysis for the four addends of the estimation error in Sections 4.2 and 4.3, the orders of the corresponding error terms are obtained. The above limit theorems (where asymptotic normality is dispensable for the following consistency result) are derived following step-by-step the same strategy as in the proof of Theorem 4.1 in Section 4.3. The discretization variances of the three multiscale estimators on the bin $[G^N_j, G^N_{j+1})$ are of order

$$\sum_{i,k\in\{1,\dots,M_N(j)\}} \frac{\alpha^{opt}_{i,M_N(j)}\alpha^{opt}_{k,M_N(j)}}{ik}\cdot i\cdot R^N_j\frac{i^2}{N^2} \sim M_N(j)\frac{R^N_j}{N^2}\sim\frac{M_N(j)}{K_N N} \ ,$$

and $M_n(j)/(nK_N)$ and $M_m(j)/(mK_N)$, respectively. For the same reason, the variances of the cross terms are of order $R_j/(NM_N(j))\sim(M_N(j)K_N)^{-1}$ and the analogous orders for the univariate estimators.

The error due to noise instead depends only on the number of observations in the considered interval. Therefore, the variance of the 'leading' addend is of order

$$\sum_{i=1}^{M_N(j)}\left(\frac{\alpha^{opt}_{i,M_m(j)}}{i}\right)^2 R^N_j \sim R^N_j/M_N^3(j)\sim\frac{N}{K_N M_N^3(j)}$$

and $M_N^{-1}(j)$ for the 'remainder' term due to end-effects for the bivariate case and analogously for the univariate estimators.
Choosing $M_N(j) \sim M_n(j) \sim M_m(j) \sim N^{1/2}K_N^{1/2}$ for every j, so that $M_N(j)N^{1/2} \to \infty$, greater than for $\widehat{\langle X,Y\rangle}_T^{multi}$, the error due to end-effects in the noise part and the discretization error dominate asymptotically the two other addends and are of order $N^{-1/4}K_N^{-1/4}$. This holds as long as $K_N N^{-1/3} \to 0$, such that $M_N(j)(N/K_N)^{-1} \to 0$ as $N \to \infty$. □

Corollary 5.1.2. *Let $S_j^{N,X}$ denote the number of observation times of $\tilde{X}$ in the bin $[(I_Y)_j^N, (I_Y)_{j+1}^N), 0 \le j \le K_N - 1$, defined through the degree of regularity of asynchronicity I_Y^N, and $S_j^{N,Y}$ the number of observations of $\tilde{Y}$ in $[(I_X)_j^N, (I_X)_{j+1}^N), 0 \le j \le K_N - 1$. The multiscale estimators*

$$\widehat{\Delta\langle X\rangle}_{(I_Y)_{j+1}^N} = \sum_{i=1}^{M_n(j)} \frac{\alpha_{i,M_n(j)}^{opt}}{i} \sum_{s=i}^{S_j^{N,X}} \left(\tilde{X}_{t_r} - \tilde{X}_{t_{r-i}}\right)^2 ,$$

$$\widehat{\Delta\langle Y\rangle}_{(I_X)_{j+1}^N} = \sum_{i=1}^{M_m(j)} \frac{\alpha_{i,M_m(j)}^{opt}}{i} \sum_{s=i}^{S_j^{N,Y}} \left(\tilde{Y}_{\tau_r} - \tilde{Y}_{\tau_{r-i}}\right)^2 .$$

weakly converge for $M_n(j) = c_n(j) \cdot \sqrt{K_N N}, M_m(j) = c_m(j) \cdot \sqrt{K_N N}$, with constants $c_n(j), c_m(j)$, to centred mixed Gaussian limits:

$$N^{1/4}K_N^{1/4}\left(\widehat{\Delta\langle X\rangle}_{(I_Y)_{j+1}^N} - \Delta\langle X\rangle_{(I_Y)_{j+1}^N}\right) \overset{st}{\rightsquigarrow} \mathbf{N}\left(0,\eta_4^2\right) , \tag{5.4a}$$

$$N^{1/4}K_N^{1/4}\left(\widehat{\Delta\langle Y\rangle}_{(I_X)_{j+1}^N} - \Delta\langle Y\rangle_{(I_X)_{j+1}^N}\right) \overset{st}{\rightsquigarrow} \mathbf{N}\left(0,\eta_5^2\right) , \tag{5.4b}$$

with almost surely finite random asymptotic variances η_4^2 and η_5^2 as $N \to \infty$. The weak convergence is stable.

Proof. For the bin-widths defined through the sequences of functions I_X^N and I_Y^N, respectively, it holds true that $0 \le \Delta(I_Y)_j^N \le T$, $0 \le \Delta(I_X)_j^N \le T$, $1 \le j \le K_N$, but the intervals are not necessarily all of order K_N^{-1}. The four uncorrelated error terms for the first estimator have variances of order

$$M_n(j)\frac{S_j^{N,X}}{N^2}, \frac{S_j^{N,X}}{M_n^3(j)}, \frac{S_j^{N,X}}{NM_n(j)}, \frac{1}{M_n(j)}$$

and analogously for the second estimator what can be deduced as in the foregoing corollary. The proof of limit theorems is derived following the proof of Theorem 4.1.
Choosing $M_n(j) \sim \sqrt{K_N N}$, $M_m(j) \sim \sqrt{K_N N}$ for all bins (with possibly j-dependent constants) each of the error terms has an asymptotic variance at most of order $K_N^{-1/2}N^{-1/2}$.
If $S_j^{N,X} = \mathcal{o}(NK_N^{-1})$, $S_j^{Y,N} = \mathcal{o}(NK_N^{-1})$, the fourth error terms of order $M_n(j)^{-1/2}$ or $M_m(j)^{-1/2}$ become the leading terms. □

Proposition 5.1.3. *The estimators* (5.2a) *and* (5.2b) *are consistent estimators for* $\int_0^T G'(t)(\sigma_t^X \sigma_t^Y \rho_t)^2\,dt$ *and* $\int_0^T G'(t)(\sigma_t^X \sigma_t^Y)^2\,dt$*, respectively, as* $K_N \to \infty$ *with* $K_N N^{-1/3} \to 0$ *as* $N \to \infty$*.*

Proof. According to Corollary 5.1.1, for the estimator (5.2a) the following asymptotic equality holds:

$$
\begin{aligned}
\hat{I}_1 &= \sum_{j=1}^{K_N} \left(\frac{\Delta\widehat{\langle X, Y\rangle}_{G_j^N}}{\Delta G_j^N} \right)^2 \frac{G^N(T)}{K_N} \\
&= \sum_{j=1}^{K_N} \left(\frac{\int_{G_{j-1}^N}^{G_j^N} \rho_t \sigma_t^X \sigma_t^Y \, dt + \mathcal{O}_p\left(N^{-1/4} K_N^{-1/4}\right)}{\Delta G_j^N} \right)^2 \frac{G^N(T)}{K_N} \\
&= \sum_{j=1}^{K_N} \left(\rho\sigma^X\sigma^Y\right)^2_{\overline{G_j^N}} \frac{G^N(T)}{K_N} + \mathcal{O}_p\left(K_N^{1/4} N^{-1/4}\right) \\
&= \sum_{j=1}^{K_N} \left(\rho\sigma^X\sigma^Y\right)^2_{G_{j-1}^N} \frac{G^N(T)}{K_N} + \mathcal{O}_p\left(K_N^{1/4} N^{-1/4}\right) \;.
\end{aligned}
$$

It has been used that $\Delta G_j^N \sim N^{-1}$ and that the multiscale estimators on disjoint bins are uncorrelated. The leading term of the stochastic error is due to the cross terms. The mean value theorem has been applied and $\overline{G_j^N}$ is some value $G_{j-1}^N \le \overline{G_j^N} \le G_j^N$. The last approximation is assured by Assumption 1, on that $\rho_t \sigma_t^X \sigma_t^Y$ is continuous and hence

$$
\begin{aligned}
&\sum_{j=1}^{K_N} \left| \left(\rho\sigma^X\sigma^Y\right)^2_{\overline{G_j^N}} - \left(\rho\sigma^X\sigma^Y\right)^2_{G_{j-1}^N} \right| \frac{G^N(T)}{K_N} \\
&\le \sup_{|t-s| \le \Delta \sup_j G_j^n} \left| \rho_t \sigma_t^X \sigma_t^Y - \rho_s \sigma_s^X \sigma_s^Y \right| G^N(T) = \mathcal{o}_{a.\,s.}(1) \;.
\end{aligned}
$$

as $K_N \to \infty$. Finally, the Riemann sum converges to the integral as $K_N \to \infty$, the stochastic error term converges to zero in probability as $N \to \infty$ with $K_N N^{-1/3} \to 0$. The last condition is due to Corollary 5.1.1 to guarantee consistency for estimating the increase of the quadratic covariation on the bins.
For the estimator (5.2b), it similarly holds true that

$$
\begin{aligned}
\hat{I}_2 &= \sum_{j=1}^{K_N} \left(\frac{\widehat{\Delta\langle X\rangle}_{G_j^N} \widehat{\Delta\langle Y\rangle}_{G_j^N}}{\left(\Delta G_j^N\right)^2} \right) \frac{G^N(T)}{K_N} \\
&= \sum_{j=1}^{K_N} \left(\frac{\int_{G_{j-1}^N}^{G_j^N} \left(\sigma_t^X\right)^2 dt \int_{G_{j-1}^N}^{G_j^N} \left(\sigma_t^Y\right)^2 dt + \mathcal{O}_p\left(N^{-1/4} K_N^{-1/4}\right)}{\Delta G_j^N} \right)^2 \frac{G^N(T)}{K_N}
\end{aligned}
$$

$$= \sum_{j=1}^{K_N} \left(\sigma^X\right)^2_{\overline{G_j^N}} \left(\sigma^Y\right)^2_{\widetilde{G_j^N}} \frac{G^N(T)}{K_N} + \mathcal{O}_p\left(K_N^{1/4} N^{-1/4}\right)$$
$$= \sum_{j=1}^{K_N} \left(\sigma^X \sigma^Y\right)^2_{G_{j-1}^N} \frac{G^N(T)}{K_N} + \mathcal{O}_p\left(K_N^{1/4} N^{-1/4}\right) .$$

To prove the last approximation using Assumption 1, we can use the same inequality as in Lemma 3.3.9:

$$\left| \left(\sigma^X\right)^2_{\overline{G_j^N}} \left(\sigma^Y\right)^2_{\widetilde{G_j^N}} - \left(\sigma^X \sigma^Y\right)^2_{G_{j-1}^N} \right|$$
$$\leq \left(\sigma^X\right)^2_{\overline{G_j^N}} \left| \left(\sigma^Y\right)^2_{\widetilde{G_j^N}} - \left(\sigma^Y\right)^2_{G_{j-1}^N} \right| + \left(\sigma^Y\right)^2_{G_{j-1}^N} \left| \left(\sigma^X\right)^2_{\overline{G_j^N}} - \left(\sigma^X\right)^2_{G_{j-1}^N} \right|$$

from which the consistency can be deduced analogously as for $\hat{I}_1$. □

Proposition 5.1.4. *The estimators* (5.2c) *and* (5.2d) *are consistent estimators for* $\int_0^T I_Y'(t)(\sigma_t^X)^2\,dt$ *and* $\int_0^T I_X'(t)(\sigma_t^Y)^2\,dt$*, respectively, as* $K_N \to \infty$ *with* $K_N N^{-1/3} \to 0$ *as* $N \to \infty$*.*

Proof. Assume that $I_X^N(T) > 0$. If $I_X^N(T) = 0$, we estimate the integral as zero. It suffices to consider the estimator (5.2c) for the first integral. The proof for the estimator (5.2d) is analogous. According to Corollary 5.1.2, for the estimator (5.2c) the following asymptotic equality holds:

$$\hat{I}_3 = \sum_{j=1}^{K_N} \left(\frac{\widehat{\Delta \langle X \rangle}_{(I_Y)_j^N}}{\Delta (I_Y)_j^N} \right) \frac{(I_Y)^N(T)}{K_N}$$
$$= \sum_{j=1}^{K_N} \left(\frac{\int_{(I_Y)_{j-1}^N}^{(I_Y)_j^N} \left(\sigma_t^X\right)^2 dt + \mathcal{O}_p\left(K_N^{-1/4} N^{-1/4}\right)}{\Delta (I_Y)_j^N} \right)^2 \frac{(I_Y)^N(T)}{K_N}$$
$$= \sum_{j=1}^{K_N} \left(\sigma^X\right)^2_{\overline{(I_Y)_j^N}} \frac{(I_Y)^N(T)}{K_N} + \mathcal{O}_p\left(K_N^{1/4} N^{-1/4}\right)$$
$$= \sum_{j=1}^{K_N} \left(\sigma^X\right)^2_{(I_Y)_{j-1}^N} \frac{(I_Y)^N(T)}{K_N} + \mathcal{O}_p\left(K_N^{1/4} N^{-1/4}\right) .$$

This equality is deduced analogously as in Proposition 5.1.3, except that the last approximation makes use of two aspects. First, since σ^X is continuous on Assumption 1 the differences of left-end points and mean values on bins converge to zero as the bin-width converges to zero. This is analogous as in Proposition 5.1.3 above. The bin-widths chosen accordingly to I_Y^N are asymptotically of order K_N^{-1} in any interval of $[0, T]$ on that the corresponding part of the integral $\int I_Y'(t)(\sigma_t^X)^2\,dt$ is strictly positive.

In any neighborhood where the limiting function I_Y is zero, bins can be of greater order than K_N^{-1} but this does not effect the approximation of the integral. Thus, as $K_N \to \infty$ with $K_N N^{-1/3} \to 0$ as $N \to \infty$, consistency of the estimator (5.2c) holds. □

Propositions 5.1.3 and 5.1.4 give the theoretical result that (5.2a)-(5.2d) provide consistent estimators for the integrals of interest. For practical implementations of the estimation method including (asymptotic) variance estimators, we still have to find rules to choose the constants that determine the multiscale frequencies on each bin and the number of bins. This is postponed to the applied part of this work in Chapter 6.
The approximation errors of the discrete Riemann sums for the corresponding integrals decrease as K_N increases, as well as the approximation errors when the mean values are replaced by the left-end points on bins, whereas the stochastic errors in the second addends increase. Thus, there is a trade-off between the two error sources. The asymptotic order of the approximation errors depends on the smoothness of ρ, σ^X and σ^Y.
In common volatility models that are used in financial econometrics, the volatilities are modeled to be Itô semimartingales again. Hence, consider the model defined by Assumption 1 with a constant correlation coefficient ρ on $[0,T]$, locally bounded drift functions, and volatility processes of the type

$$\sigma_t^X = \int_0^t u_s^X\, ds + \int_0^t v_s^X\, dB_s^X + \int_0^t w_s^X\, dB_s^{X,\perp} \;, \tag{5.5a}$$

$$\sigma_t^Y = \int_0^t u_s^Y\, ds + \int_0^t v_s^Y\, dB_s^Y + \int_0^t w_s^Y\, dB_s^{Y,\perp} \;, \tag{5.5b}$$

with continuous stochastic processes $u_s^X, u_s^Y, v_s^X, v_s^Y, w_s^X, w_s^Y$ and where $B_s^{X,\perp}, B_s^{Y,\perp}$ are Brownian motions independent of B^X and B^Y, respectively. This general stochastic volatility model, that also allows for leverage effects, includes several important volatility models as the ones by Black and Scholes [1973], Vasicek [1977], Cox et al. [1980], Heston [1993] and Chan et al. [1992]. For the model defined by Assumption 1 and (5.5a) and (5.5b), the increments of quadratic (co-)variations on intervals $[s,t]$ are of order $(t-s)$. In this particular case, the two approximation errors will both be of order K_N^{-1} and the total estimation mean square error is minimized by a choice $K_N \sim N^{1/5}$, $M_N(j) \sim N^{3/5}\ \forall j$ for estimation of the integrals with (5.2a)-(5.2d) in Propositions 5.1.3 and 5.1.4. These estimators are $N^{1/5}$-consistent in this setting.

Proposition 5.1.5 (feasible limit theorem). *The asymptotic variances* (4.8) *and* (4.10) *of the generalized multiscale estimator* (4.2) *and the one-scale subsampling estimator* (4.3) *with $M_N = c_{multi} N^{1/2}$ and $i_N = c_{sub} N^{2/3}$, can be estimated consistently by*

$$\begin{aligned}\widehat{\mathbf{AVAR}}_{multi} = &\left(c_{multi}^{-3}\left(24 + 12\,\frac{I_X^N(T) + I_Y^N(T)}{T}\right) + \frac{12}{5} c_{multi}^{-1}\right)\widehat{\eta_X^2}\widehat{\eta_Y^2} \\ &+ c_{multi}\frac{26}{35} T\left(\hat{I}_1 + \hat{I}_2\right) + c_{multi}^{-1}\frac{12}{5}\left(\widehat{\eta_Y^2}(1+\hat{I}_3) + \widehat{\eta_X^2}(1+\hat{I}_4)\right) \;, \end{aligned}\tag{5.6a}$$

$$\widehat{\mathbf{AVAR}}_{sub} = c_{sub}^{-2} 4\widehat{\eta_X^2}\widehat{\eta_Y^2} + c_{sub}\frac{2}{3}\left(\hat{I}_1 + \hat{I}_2\right) \ , \tag{5.6b}$$

where $\hat{I}_1$-$\hat{I}_4$ *are the estimators* (5.2a)-(5.2d) *and*

$$\widehat{\eta_X^2} = (2n)^{-1}\sum_{i=1}^{n}(\Delta X_{t_i})^2 \ , \ \widehat{\eta_Y^2} = (2m)^{-1}\sum_{j=1}^{m}(\Delta Y_{\tau_j})^2 \ .$$

Furthermore, the following feasible central limit theorems hold true:

$$N^{1/4}\frac{\left(\widehat{\langle X,Y\rangle}_T^{multi} - \langle X,Y\rangle_T\right)}{\widehat{\mathbf{AVAR}}_{multi}} \overset{st}{\rightsquigarrow} \mathbf{N}(0,1) \ , \tag{5.7a}$$

$$N^{1/6}\frac{\left(\widehat{\langle X,Y\rangle}_T^{sub} - \langle X,Y\rangle_T\right)}{\widehat{\mathbf{AVAR}}_{sub}} \overset{st}{\rightsquigarrow} \mathbf{N}\left(0,1\right) \ . \tag{5.7b}$$

Proof. Denote R_N^k, $k = 1,\ldots,4$, the orders of the approximation errors of the four above given integrals and their Riemann sums evaluated on the partition given K_N bins. The variance of the estimators $\widehat{\eta_X^2}$ and $\widehat{\eta_Y^2}$ for the noise variances are known to be $\mathbb{E}\left[\left(\epsilon_{t_1}^X\right)^4\right]N^{-1}$ and $\mathbb{E}\left[\left(\epsilon_{\tau_1}^Y\right)^4\right]N^{-1}$ and hence $\mathcal{O}\left(N^{-1}\right)$ on Assumption 3.
Since $I_X^N(T) \to I_X(T)$, $I_Y^N(T) \to I_Y(T)$ as $N \to \infty$, Propositions 5.1.3 and 5.1.4 yield that

$$\hat{I}_k = I_k + \mathcal{O}_p\left(R_N^k + K_N^{1/2}N^{-1/2}\right) \ , \ k = 1,\ldots,4$$

we derive that

$$\widehat{\mathbf{AVAR}}_{multi} = \mathbf{AVAR}_{multi} + \mathcal{O}_p\left(\max_k R_N^k + K_N^{1/2}N^{-1/2}\right) \ ,$$

$$\widehat{\mathbf{AVAR}}_{sub} = \mathbf{AVAR}_{sub} + \mathcal{O}_p\left(\max_k R_N^k + K_N^{1/2}N^{-1/2}\right) \ .$$

This result is obtained applying the arithmetic rules for stochastic orders stated in Proposition 1.4.2. □

For stochastic volatility models with (5.5a) and (5.5b), we conclude the following

Corollary 5.1.6. *The estimators* (5.6a) *and* (5.6b) *for the asymptotic variances of the generalized multiscale estimator* (4.2) *and the one-scale subsampling estimator* (4.3) *are* $N^{1/5}$*-consistent in the model constituted by Assumptions 1, 3, 4.3, 4.2 and equations* (5.5a) *and* (5.5b).

Remark 5.1. *For the estimation of the integrated volatility $\langle X\rangle_T$ from high-frequency observations without microstructure noise with the realized variance $\sum_{i=1}^n(\Delta X_{t_i})^2$, a consistent estimator for the asymptotic variance $2T\int_0^T G_X'(t)(\sigma_t^X)^4dt$, where G_X denotes the asymptotic quadratic variation of observation times, has been proposed in Barndorff-Nielsen and Shephard [2002] as $(2n/3)\sum_{i=1}^n(\Delta X_{t_i})^4$. In the bivariate synchronous setting one possible consistent estimator for the asymptotic variance of $\sum_{i=1}^n \Delta X_{t_i}\Delta Y_{t_i}$ is $(n/2)\sum_{i=1}^{n-1}(\Delta X_{t_i})^2\left((\Delta Y_{t_i})^2+(\Delta Y_{t_{i+1}})^2\right)$. The second addend is necessary since $n\sum_{i=1}^n(\Delta X_{t_i})^2(\Delta Y_{t_i})^2 \stackrel{p}{\to} T\int_0^T(2\rho_t^2+1)(\sigma_t^X\sigma_t^Y)^2G_{X,Y}'(t)dt$. These relations can be proved with Itô's formula and partial integration and easily comprehended by the analogy to a bivariate Gaussian distribution $(X,Y)\sim\mathbf{N}(0,\Sigma)$ with a covariance matrix Σ with entries $\sigma_X^2,\sigma_Y^2,\rho\sigma_X\sigma_Y$. Then, $\mathbb{E}X^4=3\sigma_X^4$ and $\mathbb{E}\left[X^2Y^2\right]=2\rho^2\sigma_X^2\sigma_Y^2+\sigma_X^2\sigma_Y^2$ hold true.*
If we progress to the synchronous microstructure noise setting, estimators using the techniques of Sections 2.2 and 2.3 can be found. One such method for estimation of the asymptotic variance has been provided in Christensen et al. [2010].
However, in the most general case, and even the asynchronous non-noisy setting, there is no direct extension of those estimation methods available to estimate the terms that appear in the asymptotic variance of a Hayashi-Yoshida adapted estimator as (4.2). *For that reason we have made an effort to construct the consistent histogram-based estimators* (5.2a)-(5.2d) *above.*

Finally, for the sake of completeness we state a consistent estimator for the asymptotic variance of the Hayashi-Yoshida estimator (3.2) from Theorem 3.1 in the setting of Chapter 3. Since Hayashi and Yoshida [2008] have proven a central limit theorem for the case of deterministic correlation and volatility functions, the asymptotic variance has been non-random in their setting. In a recent publication Hayashi and Yoshida [2011], in that the authors also generalize the asymptotic distribution result to a stable central limit theorem in the setting of random volatility and correlation functions, a consistent estimation method for the asymptotic variance is provided using kernel estimates. Our estimator differs from this method since we incorporate only one histogram-type estimator alike the estimator (5.2a).

Proposition 5.1.7. *Define the estimator*

$$\widehat{\mathbf{AVAR}}_{HY} := N\sum_{j=1}^{N-1}(X_{g_j}-X_{l_j})(Y_{\gamma_j}-Y_{\lambda_j})\Big[(X_{g_j}-X_{l_j})(Y_{\gamma_j}-Y_{\lambda_j}) \\ +2(X_{g_{j+1}}-X_{l_{j+1}})(Y_{\gamma_{j+1}}-Y_{\lambda_{j+1}})\Big]-3T\tilde I_1$$

with

$$\tilde I_1 := \sum_{j=1}^{K_N}\left(\frac{\Delta\widehat{\langle X,Y\rangle}^{HY}_{G_j^N}}{\Delta G_j^N}\right)^2\frac{G^N(T)}{K_N}$$

being the histogram-based estimator for $\int_0^T(\rho_t\sigma_t^X\sigma_t^Y)^2G'(t)dt$, similarly to (5.2a) *above.*

If the continuous semimartingale is not latent but observable, we replace the multiscale estimators on bins by Hayashi-Yoshida estimators of the type

$$\Delta\widehat{\langle X,Y\rangle}^{HY}_{G_j^N} := \sum_{r\in[G_j^N,G_{j+1}^N)} (X_{g_r}-X_{l_r})(Y_{\gamma_r}-Y_{\lambda_r}) \;.$$

It holds true that

$$\widehat{\mathbf{AVAR}}_{HY} \xrightarrow{p} \mathbf{AVAR}_{HY} =$$
$$T\int_0^T G'(t)\left(\sigma_t^X\sigma_t^Y\right)^2\left(\rho_t^2+1\right)dt + T\int_0^T\left(F'(t)\left(\sigma_t^X\sigma_t^Y\right)^2 dt + 2H'(t)\left(\rho_t\sigma_t^X\sigma_t^Y\right)^2 dt\right)$$

on the Assumptions 1, 2(a) and 3.1. Thus, we have on hand a consistent estimator for the asymptotic variance of the Hayashi-Yoshida estimator and the feasible stable central limit theorem

$$\frac{\widehat{\langle X,Y\rangle}_T^{(HY)}}{\sqrt{\widehat{\mathbf{AVAR}}_{HY}}} \overset{st}{\rightsquigarrow} \mathbf{N}(0,1) \;.$$

Proof. The proof will be divided into three parts in that the sum of squared products, products of consecutive increments and the histogram estimator are considered, respectively. Denote $X_j^+ = X_{g_j} - X_{T_j}$, $X_j^- = X_{T_{j-1}} - X_{l_j}$ and $X_j^S = X_{T_j} - X_{T_{j-1}}$, $j = 1,\ldots,N$. In the first step it is proved that

$$N\sum_{j=1}^{N-1}\left(X_{g_j}-X_{l_j}\right)^2\left(Y_{\gamma_j}-Y_{\lambda_j}\right)^2 \xrightarrow{p} T\int_0^T G'(t)\left(\sigma_t^X\sigma_t^Y\right)^2\left(2\rho_t^2+F'(t)\right)dt \;.$$

$$N\sum_{j=1}^{N-1}\left(X_j^+ + X_j^S + X_j^-\right)^2\left(Y_j^+ + Y_j^S + Y_j^-\right)^2 = N\sum_{j=1}^{N-1}\Big((X_j^+)^2(Y_j^S+Y_j^-)^2$$
$$+(Y_j^+)^2(X_j^S+X_j^-)^2 + (X_j^-)^2(Y_j^S)^2 + (Y_j^-)^2(X_j^S)^2 + (X_j^SY_j^S)^2\Big) + o_p(1)$$

All centred addends have a variance tending to zero as $N\to\infty$ and converge to zero in probability. The sum of the first four addends times the factor N/T has been proved to converge in probability to $\int_0^T F'(t)(\sigma_t^X\sigma_t^Y)^2dt$ in Lemma 3.3.7 where this term has appeared in the sequence of conditional variances of the error due to non-synchronicity. Hence, it remains to prove that $N\sum_{j=1}^{N-1}(X_j^SY_j^S)^2 \xrightarrow{p} T\int_0^T(2\rho_t^2+1)(\sigma_t^X\sigma_t^Y)^2G'(t)dt$. For this purpose recall the notation from Section 3.3 in the proof of Proposition 3.3.1 after a measure change such that the drift terms are zero. With $L_t = \int_0^t \sigma_s^X dW_s^X$, $M_t = \int_0^t \sigma_s^Y dW_s^Y$, $L_i = L_{T_i}, M_i = M_{T_i}$, we can write the term

$$N\sum_{j=1}^{N-1}((L-L_{i-1})_{T_i}(M-M_{i-1})_{T_i})^2 = N\sum_{i=1}^{N-1}\Big(2\int_0^{T_i}(L-L_{i-1})_t(M-M_{i-1})_t^2 d(L-L_{i-1})_t$$

$$
\begin{aligned}
&+2\int_0^{T_i}(L-L_{i-1})_t^2(M-M_{i-1})_t d(M-M_{i-1})_t\\
&+4\int_0^{T_i}(L-L_{i-1})_t(M-M_{i-1})_t d\langle M-M_{i-1},L-L_{i-1}\rangle_t\\
&+\int_0^{T_i}(M-M_{i-1})_t^2 d\langle M-M_{i-1}\rangle_t+\int_0^{T_i}(L-L_{i-1})_t^2 d\langle L-L_{i-1}\rangle_t\Bigg)\ ,
\end{aligned}
$$

where we have applied Itô's formula. The sum of the first two addends converges to zero in probability since it is centred and the variance converges to zero. Since

$$
\int_0^{T_i}(L-L_{i-1})_t(M-M_{i-1})_t d\langle M-M_{i-1},L-L_{i-1}\rangle_t=\int_{T_{i-1}}^{T_i}(L-L_{i-1})_t(M-M_{i-1})_t d\langle M,L\rangle_t,
$$

the sum of the third addends has been considered in the proof of Proposition 3.3.2 as part of the quadratic variation of the discretization error of the closest synchronous approximation and converges in probability to $2T\int_0^T G'(t)(\rho_t\sigma_t^X\sigma_t^Y)^2dt$. The remaining sum of the fourth addends is also similar to the other part of the quadratic variation in the proof of Proposition 3.3.2. With an analogous approximation and integration by parts one obtains

$$
\begin{aligned}
&\int_{T_{i-1}}^{T_i}(M-M_{i-1})_t^2 d\langle M\rangle_t+\int_{T_{i-1}}^{T_i}(L-L_{i-1})_t^2 d\langle L\rangle_t\\
&=\int_{T_{i-1}}^{T_i}\langle M-M_{i-1}\rangle_t d\langle M\rangle_t+\int_{T_{i-1}}^{T_i}\langle L-L_{i-1}\rangle_t d\langle L\rangle_t+\mathcal{O}_p(1)\\
&=\int_{T_{i-1}}^{T_i}d(\langle L-L_{i-1}\rangle_t\langle M-M_{i-1}\rangle_t)+\mathcal{O}_p(1)
\end{aligned}
$$

and the convergence of the above given term to $T\int_0^T(\sigma_t^X\sigma_t^Y)^2G'(t)dt$.
In the second part of the proof we are concerned with the term

$$
\begin{aligned}
&2N\sum_{j=1}^{N-1}(X_j^+ + X_j^S + X_j^-)(X_{j+1}^+ + X_{j+1}^S + X_{j+1}^-)(Y_j^+ + Y_j^S + Y_j^-)(Y_{j+1}^+ + Y_{j+1}^S + Y_{j+1}^-)\\
&\quad=2N\sum_{j=1}^{N-1}\left(X_j^SY_j^SX_{j+1}^SY_{j+1}^S + X_j^+Y_j^SX_{j+1}^-Y_{j+1}^S + Y_j^+X_j^SY_{j+1}^-X_{j+1}^S\right)+\mathcal{O}_p(1)\ .
\end{aligned}
$$

The sum incorporating all centred addends converges to zero in probability. The last two addends capture the only dependence between consecutive addends in the error due to non-synchronicity (3.6), namely when next-tick interpolations and previous-tick interpolations at the same $T_i, i=1,\ldots,N$ are included. Those have appeared in the proof of Lemma 3.3.7 and have been proved to converge to $T\int_0^T 2H'(t)(\sigma_t^X\sigma_t^Y)^2dt$ in probability. That $2N\sum(X_j^SY_j^SX_{j+1}^SY_{j+1}^S)\xrightarrow{p}2\int_0^T G'(t)(\rho_t\sigma_t^X\sigma_t^Y)^2dt$ follows with the methodology of Lemma 4.3.10 and the concept of a time-change in the asymptotic quadratic variation of refresh times such that (4.13) holds true. Using the mean value theorem and $(\Delta T_j)^2 - \Delta T_j \Delta T_{j+1} = \Delta T_j(\Delta T_j - T/N) + \Delta T_j(T/N - \Delta T_{j+1})$ together with the Cauchy-Schwarz inequality

$$N\left|\sum_{j=1}^{N-1}\left(\Delta T_j\left(\Delta T_j - \frac{T}{N}\right)\right)\right| \leq N\sqrt{\sum_{j=1}^{N-1}(\Delta T_j)^2}\sqrt{\sum_{j=1}^{N-1}\left(\Delta T_j - \frac{T}{N}\right)^2}$$

yields the result.
The Hayashi-Yoshida estimators on the bins in the above given histogram-based estimator fulfill

$$\Delta\widehat{\langle X,Y\rangle}^{(HY)}_{G_j^N} = \int_{G_{j-1}^N}^{G_j^N} \rho_t \sigma_t^X \sigma_t^Y \, dt + \mathcal{O}_p\left(K_N^{1/2} N^{-1/2}\right)$$

so that the estimation error of the sum is of order $K_N^{3/4} N^{-1/2}$ in probability and for $K_N \to \infty$, $N \to \infty$, $K_N N^{-2/3} \to 0$ consistency holds and we conclude consistency of the estimator of the asymptotic variance. □

5.2 Independent Poisson observation schemes

In this section, we consider the model in which the sequences of observation times are supposed to be realizations of two homogeneous Poisson processes that are mutually independent and independent of the processes $\tilde{X}$ and $\tilde{Y}$. Although this model can be criticized for its flaw that sampling schemes of two correlated processes follow two independent processes and time homogeneity, what might seem to be unrealistic in financial applications, independent and homogeneous Poisson sampling times designs constitute the most commonly used model in this research area (cf. Zhang [2006], Hayashi and Yoshida [2005] among others).
Let $\tilde{n}^{(n)}(t)$ and $\tilde{m}^{(n)}(t)$ be sequences of two independent homogeneous Poisson processes with parameters Tn/θ_1 and Tn/θ_2 ($n \in \mathbb{N}$), such that the waiting times between jumps of $\tilde{n}^{(n)}$ and $\tilde{m}^{(n)}$ are exponentially distributed with expectations $\mathbb{E}\left[\Delta t_i^{(n)}\right] = \theta_1/n$ and $\mathbb{E}\left[\Delta \tau_j^{(n)}\right] = \theta_2/n$, $i \in \mathbb{N}, j \in \mathbb{N}$. Thus, $\tilde{n}^{(n)}(T)$ and $\tilde{m}^{(n)}(T)$ correspond to the sequences giving the numbers of observation times of $\tilde{X}$ and $\tilde{Y}$ in the time span $[0,T]$. The increments of the sampling times of the closest synchronous approximation (3.4) correspond to the maxima of the exponentially distributed waiting times:

$$\Delta T_k^{(n)} \sim F(t) = 1 - \exp\left(-\frac{tn}{\theta_1}\right) - \exp\left(-\frac{tn}{\theta_2}\right) + \exp\left(-tn\left(\frac{1}{\theta_1}+\frac{1}{\theta_2}\right)\right), k \in \mathbb{N}\,.$$

Denote $\tilde{N}(T)^{(n)} = \max_{N\in\mathbb{N}}\{\sum_{k=0}^{N}\Delta T_k^{(n)} \leq T\}$. We focus on the characteristics of the sampling schemes affecting the asymptotics of both, the synchronized realized covariance estimator (3.2) from Chapter 3 and the generalized multiscale estimator (4.2) from

Chapter 4. Namely those quantities of interest are the quadratic (co-)variations of times defined in Definition 3.2.3 and the degree of regularity of non-synchronicity from Definition 4.2.1.

Proposition 5.2.1. *In the independent homogeneous Poisson model, it holds true that*

$$G^N(t) \xrightarrow{p} 2\left(1-\frac{2\theta_1^2\theta_2^2}{\theta_1^2\theta_2^2+(\theta_1^2+\theta_2^2)(\theta_1+\theta_2)^2}\right)\frac{t}{T}\quad\left(=\frac{14}{9}\frac{t}{T}\quad \textit{if}\ \ \theta_1=\theta_2=\theta\right)\ , \tag{5.8a}$$

$$F^N(t) \xrightarrow{p} \left(\frac{2\theta_1\theta_2}{(\theta_1^2+\theta_1\theta_2+\theta_2^2)}+\frac{4\theta_1^2\theta_2^2}{\left(\theta_1+\theta_2-\frac{\theta_1\theta_2}{\theta_1+\theta_2}\right)^2(\theta_1+\theta_2)^2}\right)\frac{t}{T} \tag{5.8b}$$
$$\left(=\frac{10}{9}\frac{t}{T}\quad \textit{if}\ \ \theta_1=\theta_2=\theta\right)\ ,$$

$$H^N(t) \xrightarrow{p} 2\left(\frac{1}{\left(\theta_1+\theta_2-\frac{\theta_1\theta_2}{\theta_1+\theta_2}\right)^2}\frac{\theta_1^2\theta_2^2}{(\theta_1+\theta_2)^2}\right)\frac{t}{T}\quad\left(=\frac{2}{9}\frac{t}{T}\quad \textit{if}\ \ \theta_1=\theta_2=\theta\right)\ , \tag{5.8c}$$

$$I_X^N(t) \xrightarrow{p} \frac{\theta_1\theta_2 t}{(\theta_1+\theta_2)^2}\ ,\ I_Y^N(t) \xrightarrow{p} \frac{\theta_1\theta_2 t}{(\theta_1+\theta_2)^2}\quad\left(=\frac{z\,t}{(z+1)^2}\quad \textit{if}\ \ \theta_1=z\theta_2\right)\ . \tag{5.8d}$$

Proof. We make use of the basic properties of mutually independent homogeneous Poisson processes in this proof. Those are Markovian and the exponential distributions of the increments between arrival times are memoryless. Wald's identity ensures that $\mathbb{E}\left[\sum_{k=0}^{\tilde N(T)^{(n)}}\Delta T_k^{(n)}\right]=\mathbb{E}\left[\tilde N(T)^{(n)}\right]\mathbb{E}\left[\Delta T_1^{(n)}\right]$. For the proofs of these results on Poisson processes and further information we refer to Cox and Isham [1980].
First of all we ascertain that $t_i^{(n)}\neq \tau_j^{(n)}\,\forall\,(i,j)\in\{1,\ldots,\tilde n^{(n)}(T)\}\times\{1,\ldots,\tilde m^{(n)}(T)\}$ almost surely. For arbitrarily fixed i, the expected values of next-tick, previous-tick and refresh time instants yield

$$\mathbb{E}\left[g_i^{(n)}-T_i^{(n)}\right]=\mathbb{E}\left[\left(g_i^{(n)}-T_i^{(n)}\right)\mid T_i^{(n)}=\gamma_i^{(n)}\right]\mathbb{P}\left(T_i^{(n)}=\gamma_i^{(n)}\right)=\frac{\theta_1}{n}\frac{\theta_2}{\theta_1+\theta_2}\ ,$$

$$\mathbb{E}\left[\gamma_i^{(n)}-T_i^{(n)}\right]=\frac{\theta_2}{n}\frac{\theta_1}{\theta_1+\theta_2}\ ,$$

$$\mathbb{E}\left[T_i^{(n)}-l_{i+1}^{(n)}\right]=\int_0^\infty y\frac{n}{\theta_2}e^{-\frac{yn}{\theta_2}}e^{-\frac{yn}{\theta_1}}\,dy=\frac{1}{n}\frac{\theta_1^2\theta_2}{(\theta_1+\theta_2)^2}\ ,$$

$$\mathbb{E}\left[T_i^{(n)}-\lambda_{i+1}^{(n)}\right]=\frac{1}{n}\frac{\theta_1\theta_2^2}{(\theta_1+\theta_2)^2}\ ,$$

$$\mathbb{E}\left[T_{i+1}^{(n)}-T_i^{(n)}\right]=\frac{\theta_1}{n}+\frac{\theta_2}{n}-\frac{1}{n}\frac{\theta_1\theta_2}{\theta_1+\theta_2}\ .$$

The conditional expectations given that the ith refresh time $T_i^{(n)}=\gamma_i^{(n)}$ is an arrival time of $\tilde{m}^{(n)}$ yield $\mathbb{E}\left[T_{i+1}^{(n)}-T_i^{(n)}|T_i^{(n)}=\gamma_i^{(n)}\right]=\mathbb{E}\left[T_{i+1}^{(n)}-T_i^{(n)}\right]$ and

$$\mathbb{E}\left[T_i^{(n)}-l_{i+1}^{(n)}|T_i^{(n)}=\gamma_i^{(n)}\right]=\mathbb{E}\left[T_i^{(n)}-l_{i+1}^{(n)}\right]\ ,$$

since the latter previous-tick interpolation is zero with probability 1 if $T_i^{(n)}\neq\gamma_i^{(n)}$. Only for $(T_i^{(n)}-\lambda_i^{(n)})$ the conditional expectation differs from the unconditional and can be calculated by further conditioning

$$\begin{aligned}
&\mathbb{E}\left[T_i^{(n)}-\lambda_i^{(n)}|T_i^{(n)}=\gamma_i^{(n)}\right]=\\
&\quad\mathbb{E}\left[T_i^{(n)}-\lambda_i^{(n)}|T_i^{(n)}=\gamma_i^{(n)},\,T_{i-1}^{(n)}=\lambda_i^{(n)}\right]\mathbb{P}\left(T_{i-1}^{(n)}=\lambda_i^{(n)}|T_i^{(n)}=\gamma_i^{(n)}\right)\\
&\quad+\mathbb{E}\left[T_i^{(n)}-\lambda_i^{(n)}|T_i^{(n)}=\gamma_i^{(n)},\,T_{i-1}^{(n)}=l_i^{(n)}\right]\mathbb{P}\left(T_{i-1}^{(n)}=l_i^{(n)}|T_i^{(n)}=\gamma_i^{(n)}\right)\\
&=\left(\theta_1+\theta_2-\frac{\theta_1\theta_2}{\theta_1+\theta_2}\right)\frac{\theta_1}{\theta_1+\theta_2}+2\theta_1\frac{\theta_2}{\theta_1+\theta_2}\ ,
\end{aligned}$$

where the factor $2\theta_1$ in the second addend is simply the expectation of the waiting time for two jumps of $\tilde{n}$. Here, we have used some simplifying symmetry aspects, a rigorous proof using the density functions is obtained by calculation of

$$\begin{aligned}
\mathbb{E}\left[T_i^{(n)}-\lambda_i^{(n)}\mathbb{1}_{\{T_i^{(n)}=\gamma_i^{(n)},T_{i-1}^{(n)}=\lambda_i^{(n)}\}}\right]&=\int_0^\infty\int_x^\infty x\frac{n}{\theta_1}e^{-x\frac{n}{\theta_1}}e^{-y\frac{n}{\theta_2}}y\frac{n}{\theta_2}e^{-x\frac{n}{\theta_1}}e^{-y\frac{n}{\theta_2}}\,dx\,dy\\
&=\frac{2\theta_1\theta_2}{\theta_1+\theta_2}\ .
\end{aligned}$$

The conditional expectations on $T_i^{(n)}=g_i^{(n)}$ are deduced analogously. Since

$$\mathbb{E}\left[T_i^{(n)}-l_i^{(n)}\right]=\mathbb{E}\left[T_i^{(n)}-T_{i-1}^{(n)}\right]+\mathbb{E}\left[T_{i-1}^{(n)}-l_i^{(n)}\right]$$

and the (conditional) expectations of the products occurring in G^N,F^N,H^N equal the products of (conditional) expectations thanks to the memorylessness of exponential distributions, the latter results suffice to apply the law of large numbers to the empirical (co-)variations of times. For the asymptotics of $G^N(T),F^N(T)$ and $H^N(T)$, we conclude for the number of addends $\tilde{N}(T)^{(n)}$, that $\mathbb{E}\tilde{N}(T)^{(n)}=(T/\theta)n+\mathcal{O}(n)$ with $\theta=\theta_1+\theta_2-(\theta_1\theta_2)/(\theta_1+\theta_2)$ what follows from $\mathbb{E}\tilde{N}(T)^{(n)}\mathbb{E}\left[\Delta T_1^{(N)}\right]=T+\mathcal{O}_p(n^{-1})$ and

$\mathbb{V}\mathrm{ar}\left(\tilde{N}(T)^{(n)}\right) = \mathcal{O}(n^{-1})$ since

$$\mathbb{V}\mathrm{ar}\left(\sum_{k=0}^{\tilde{N}(T)^{(n)}} \Delta T_k^{(n)}\right) = \mathbb{V}\mathrm{ar}\left(\tilde{N}(T)^{(n)}\right)\mathbb{E}\left[\left(\Delta T_1^{(n)}\right)^2\right] + \mathbb{E}\left[\tilde{N}(T)^{(n)}\right]\mathbb{V}\mathrm{ar}\left(\Delta T_1^{(n)}\right) \ .$$

The exact probability mass functions of the counting processes $\tilde{N}(t)^{(n)}$ associated with the maxima of the waiting times $\Delta t_i^{(n)}, \Delta\tau_j^{(n)}$ have a quite complicated form, so that we only give the last two results on the expectation and the variance that are necessary for the proof of the proposition.
From the preceding conclusions, it follows that

$$G^N(t) = \frac{\tilde{N}(T)^{(n)}}{T}\sum_{T_i^{(n)}\le t}\left(\Delta T_i^{(n)}\right)^2 \xrightarrow{p} \frac{n^2}{\theta^2}\left(\frac{2\theta_1^2}{n^2} + \frac{2\theta_2^2}{n^2} - 2\left(\frac{\theta_1\theta_2}{(\theta_1+\theta_2)}\right)^2\frac{1}{n^2}\right)\frac{t}{T} \ ,$$

$$\begin{aligned}F^N(t) = \frac{\tilde{N}(T)^{(n)}}{T}\sum_{T_{i+1}^{(n)}\le t} &(T_i^{(n)} - \lambda_i^{(n)})(g_i^{(n)} - T_i^{(n)}) + \left(T_i^{(n)} - l_i^{(n)}\right)\left(\gamma_i^{(n)} - T_i^{(n)}\right)\\
&+ \Delta T_{i+1}^{(n)}\left(T_i^{(n)} - l_{i+1}^{(n)}\right) + \Delta T_{i+1}^{(n)}\left(T_i^{(n)} - \lambda_{i+1}^{(n)}\right)\\
\xrightarrow{p} \frac{t}{T\theta^2}&\left(\frac{\theta_1\theta_2}{(\theta_1+\theta_2)}\left(2\theta_1 + 2\theta_2 - 2\frac{\theta_1\theta_2}{(\theta_1+\theta_2)} + \frac{2\theta_1\theta_2}{(\theta_1+\theta_2)}\right)\right.\\
&\left. + \left(\theta_1+\theta_2 - \frac{\theta_1\theta_2}{(\theta_1+\theta_2)}\right)\frac{\theta_1^2\theta_2 + \theta_1\theta_2^2}{(\theta_1+\theta_2)^2}\right) \ ,\end{aligned}$$

$$\begin{aligned}H^N(t) = \frac{\tilde{N}(T)^{(n)}}{T}&\sum_{T_{i+1}^{(n)}\le t}\left(T_i^{(n)} - l_{i+1}^{(n)}\right)\left(g_i^{(n)} - T_i^{(n)}\right) + \left(T_i^{(n)} - \lambda_{i+1}^{(n)}\right)\left(\gamma_i^{(n)} - T_i^{(n)}\right)\\
\xrightarrow{p} &\frac{t}{T\theta^2}\frac{\theta_1^2\theta_2^2(\theta_1+\theta_2)}{(\theta_1+\theta_2)^3} \ .\end{aligned}$$

Inserting θ we obtain formulae (5.8a)-(5.8c). In the evaluation of G^N we have also used the second moment of $\Delta T_1^{(n)}$ which can be calculated using the above given distribution function.
Considering the degree of regularity of non-synchronicity defined in Definition 4.2.1 we have learned in Sections 4.2 and 4.3 that it is due to observation time aggregations distributed according to case 2 (cf. Section 4.2) where two jumps of the same process occur in a time interval in that the other process has no jumps. The probability that this is the case for the ith step when applying Algorithm 3.1 equals

$$\mathbb{P}\left(g_{i+1}^{(n)} = g_i^{(n)}\right) = \mathbb{P}\left(g_i^{(n)} > \gamma_i^{(n)}\,,\, g_i^{(n)} \ge \gamma_{i,+}^{(n)}\right)$$

$$= \mathbb{P}\left(g_i^{(n)} \geq \gamma_{i,+}^{(n)} \,|\, g_i^{(n)} > \gamma_i^{(n)}\right) \mathbb{P}\left(g_i^{(n)} > \gamma_i^{(n)}\right)$$
$$= \frac{\theta_2}{\theta_1+\theta_2}\frac{\theta_1}{\theta_1+\theta_2} = \frac{\theta_1\theta_2}{(\theta_1+\theta_2)^2}$$

for every $i \in \{1, \ldots, \tilde{N}(T)^{(n)}\}$ and analogously

$$\mathbb{P}\left(\gamma_{i+1}^{(n)} = \gamma_i^{(n)}\right) = \frac{\theta_1\theta_2}{(\theta_1+\theta_2)^2} \; .$$

Recall that for almost surely totally disjoint sets of arrival times of $\tilde{n}$ and $\tilde{m}$ these probabilities have to be equal. With the law of large numbers

$$I_X^N(t) = \frac{T}{\tilde{N}(T)^{(n)}} \sum_{g_j^{(n)} \leq t} \mathbb{1}_{\{g_j^{(n)} = g_{j-1}^{(n)}\}} \xrightarrow{p} \frac{\theta_1\theta_2 t}{(\theta_1+\theta_2)^2} \; ,$$

$$I_Y^N(t) = \frac{T}{\tilde{N}(T)^{(n)}} \sum_{\gamma_j^{(n)} \leq t} \mathbb{1}_{\{\gamma_j^{(n)} = \gamma_{j-1}^{(n)}\}} \xrightarrow{p} \frac{\theta_1\theta_2 t}{(\theta_1+\theta_2)^2} \; .$$

□

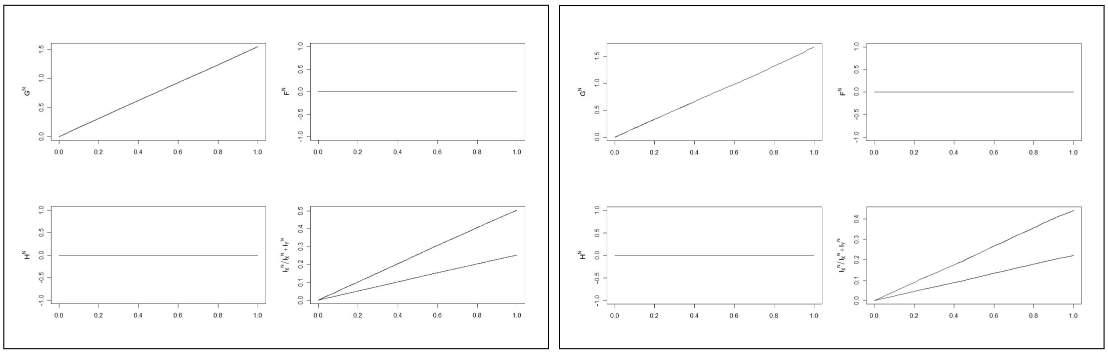


Figure 5.1: Quadratic (Co-)variations of times for homogeneous Poisson sampling.

Figure 5.1 shows the quadratic (co-)variations of times and degrees of regularity of non-synchronicity for simulated mutually independent homogeneous Poisson processes. On the left-hand side both parameters have been set $\theta = 1$ for $T = 1$ and $n = 30000$. The limits are linear increasing functions on $[0, 1]$ with slope $14/9, 10/9, 2/9$ and $1/4$, respectively. On the right-hand side we see the (co-)variations of times and degrees of regularity of asynchronicity for $T = 1, n = 30000, \theta_1 = 1, \theta_2 = 0.5$. Those tend to linear limiting functions with slope $82/49, 44/49, 8/49$ and $2/9$, respectively.
In the model of non-synchronously observed Itô processes X and Y as considered in Chapter 3 and observation times following an independent Poisson sampling scheme of the above given form, we derive the following stable central limit theorem as special case

of Theorem 3.1:

Corollary 5.2.2. *The estimation error of the synchronized realized covariance estimator* (3.2) *converges on the Assumption 1 conditionally on the independent Poisson sampling scheme with* $0 < \theta_1 < \infty$ *and* $0 < \theta_2 < \infty$ *stably in law to a centred mixed Gaussian distribution:*

$$\sqrt{\tilde{N}(T)^{(n)}}\left(\sum_{i=0}^{\tilde{N}(T)^{(n)}}\left(X_{g_i^{(n)}} - X_{l_i^{(n)}}\right)\left(Y_{\gamma_i^{(n)}} - Y_{\lambda_i^{(n)}}\right) - \langle X\,,\,Y\rangle_T\right) \overset{st}{\rightsquigarrow} \mathbf{N}\left(0\,,\,v_T\right)\;, \quad (5.9)$$

with the asymptotic variance

$$v_T = 2\int_0^T\left(\rho_t\sigma_t^X\sigma_t^Y\right)^2 dt + \left(2\frac{\theta_1\theta_2}{\theta(\theta_1+\theta_2)}+1\right)\int_0^T\left(\sigma_t^X\sigma_t^Y\right)^2$$

where the two addends come from the asymptotic variances of the discretization error D_T^N *of the closest synchronous approximation* (3.5) *and the additional error* A_T^N *due to interpolations* (3.6)*, respectively, and* $\theta = \theta_1 + \theta_2 - \frac{\theta_1\theta_2}{\theta_1+\theta_2}$.

Proof. It is a basic result in the theory of extreme values that for the supremum of n i. i. d. exponentially distributed waiting times ΔT_i with $\mathbb{E}\Delta T_i = n^{-1}$, it holds true that $\sup_i(\Delta T_i) = \mathcal{O}_p(\log(n)/n)$. We refer to de Haan and Ferreira [2006] for a proof. In the setting of mutually independent homogeneous Poisson processes with parameters Tn/θ_1 and Tn/θ_2, we conclude that $\sup_{i\in\{1,\dots,\tilde{N}(T)^{(n)}\}} = \mathcal{O}_p\left(\log\tilde{N}(T)^{(n)}/\tilde{N}(T)^{(n)}\right)$. Hence, Assumption 2(a) holds for the sampling design where the order in condition (a) holds in probability. Then all findings in the proofs of Propositions 3.3.2 and 3.3.5 stay valid when we insert the (co-)variations of time deduced above in the limits of the variances. □

The stable convergence holds conditionally given the observation times (cf. the discussion following Assumption 2), what means that endogenous observation times are not covered but Poisson sampling independent to the processes $\tilde{X}$ and $\tilde{Y}$.
The asymptotic variance of the mixed Gaussian limit is in line with the result by Hayashi and Yoshida [2008] and Hayashi and Yoshida [2011]. We remark that one has to pay attention to the proportionality to θ in the rate $\tilde{N}(T)^{(n)}$ when comparing the asymptotic variances to the one in Hayashi and Yoshida [2011].
The following versions of the stable central limit theorems from Theorem 4.1 and Corollary 4.2.2 complete our analysis of the homogeneous Poisson sampling setting.

Corollary 5.2.3. *On the Assumptions 1 and 3, the generalized multiscale estimator* (4.2) *with noise-optimal weights* (4.14)*, and* $M_N = c_{multi}\cdot\sqrt{N}$*, converges conditionally on the independent Poisson sampling scheme with* $0 < \theta_1 < \infty$ *and* $0 < \theta_2 < \infty$ *stably in law with rate* $N^{1/4}$ *to a mixed normal limit:*

$$N^{1/4}\left(\widehat{\langle X, Y\rangle}_T^{multi} - \langle X, Y\rangle_T\right) \overset{st}{\rightsquigarrow} \mathbf{N}\left(0, \mathbf{AVAR}_{multi}^{poiss}\right)$$

with the asymptotic variance

$$
\begin{aligned}
\mathbf{AVAR}_{multi}^{poiss} &= c_{multi}^{-3}\left(24 + 12\,\frac{2\theta_1\theta_2}{(\theta_1+\theta_2)^2}\right)\eta_X^2\eta_Y^2 + c_{multi}^{-1}\,\frac{12\eta_X^2\eta_Y^2}{5} \\
&+ c_{multi}\,\frac{26}{35}\int_0^T 2\left(1 - \frac{2\theta_1^2\theta_2^2}{\theta_1^2\theta_2^2 + (\theta_1^2+\theta_2^2)(\theta_1+\theta_2)^2}\right)(\sigma_t^X\sigma_t^Y)^2(1+\rho_t^2)\,dt \qquad (5.10) \\
&+ c_{multi}^{-1}\frac{12}{5}\left(\eta_Y^2\int_0^T(1+\frac{\theta_1\theta_2}{\theta_1+\theta_2})(\sigma_t^X)^2\,dt + \eta_X^2\int_0^T(1+\frac{\theta_1\theta_2}{\theta_1+\theta_2})(\sigma_t^Y)^2\,dt\right)\,.
\end{aligned}
$$

On the same Assumptions, the one-scale subsampling estimator with subsampling frequency $i_N = c_{sub}\cdot N^{2/3}$ *converges conditionally on the sampling scheme stably in law with rate* $N^{1/6}$ *to a mixed Gaussian limiting distribution:*

$$
N^{1/6}\left(\widehat{\langle X,Y\rangle}_T^{sub} - \langle X,Y\rangle_T\right) \stackrel{st}{\rightsquigarrow} \mathbf{N}\left(0, \mathbf{AVAR}_{sub}^{poiss}\right)\ ,
$$

with the asymptotic variance

$$
\begin{aligned}
\mathbf{AVAR}_{sub}^{poiss} &= c_{sub}^{-2}\,4\eta_X^2\eta_Y^2 \qquad (5.11) \\
&+ c_{sub}\,\frac{2}{3}\int_0^T 2\left(1 - \frac{2\theta_1^2\theta_2^2}{\theta_1^2\theta_2^2 + (\theta_1^2+\theta_2^2)(\theta_1+\theta_2)^2}\right)(\sigma_t^X\sigma_t^Y)^2(1+\rho_t^2)\,dt\ .
\end{aligned}
$$

Proof. Since Assumption 2(b) also holds for the sampling design when the order in condition (b) is in probability, the proofs of Theorem 4.1 and Corollary 4.2.2 stay valid with the according asymptotic degrees of regularities and asymptotic quadratic variation of time for the closest synchronous approximation derived in this section above. □

5.3 Sample size dependent noise variances and relaxing some assumptions

The theoretical results in Chapter 4 are deduced on the Assumptions 1, that the hidden efficient processes follow the dynamics of continuous semimartingales with continuous volatility processes and locally bounded drift processes, and Assumption 2 that on a fixed time span $[0,T]$ the number of observations $n \sim m \to \infty$ and the suprema of time instants tend to zero sufficiently fast. Assumption 3 imposes the condition that the additive noise processes are centred i. i. d. , have finite fourth moments and are mutually independent and independent of each other. These assumptions are convenient to facilitate a comprehensible analysis and acceptable readability. We keep to the cornerstone of a latent semimartingale model with an independent additive noise component. However, some regularity assumptions have been chosen too restrictively, since they can be relaxed such that the estimation approach proposed in Chapter 4 will stay valid without any further adjustment. Another aspect that urges us to give some model extensions arises from an applications angle. Therefore, we list some important possible model specifications and their influence on the estimation approach in the following.

- **N-dependent noise variances**
  The microstructure processes are discrete-time processes that occur as observation noise at the sampling times $t_i^{(n)}$, $i = 0, \ldots, n$ and $\tau_j^{(m)}$, $j = 0, \ldots, m$. So far we have considered i. i. d. noise with
  $$\mathbb{E}\left[\epsilon^X_{t_i^{(n)}}\right] = 0 \text{ and } \mathbb{E}\left[\left(\epsilon^X_{t_i^{(n)}}\right)^2\right] = \eta_X^2 \;,\; 0 \le i \le n \;,$$
  and analogous for $\tilde{Y}$, where the distribution of the observation errors does not at all depend on the number $(n+1)$ or $(m+1)$ of observations.
  One could be interested in the case where the noise level may vary with $N \sim n \sim m$. This was already included in the analysis of a diffusion with Gaussian noise considered in Gloter and Jacod [2001]. The primary motivation to accommodate dependence of the noise on the sample size in the model originates from the economic background. In empirical studies of (ultra) high-frequency financial data that inspired both, economists and statisticians, to analyze latent semimartingale models with additive noise, two aspects of market microstructure frictions are reported. First, the realized variance increases as the sampling frequency increases as can be seen in the signature plot in Figure 0.1. At the same time a positive correlation between the absolute value of the observation error and the time interval to the previous observation time is significantly present. This and other influences suggest to rather model the observed log-prices as sum of a latent semimartingale and noise for that the variance decreases in N as reported in Kalnina and Linton [2008] and Awartani et al. [2009], among others.
  Hence, from an applied point of view it is desirable that an estimation method is practicable in that setting. We show in the following that this is the case for the generalized multiscale estimator (4.2). We benefit from the effort of carrying out the theory for a sophisticated strategy to deal with non-synchronicity also for the model including observation noise. If an estimation approach uses previous-tick interpolations, as the one proposed in Barndorff-Nielsen et al. [2008b], these methods are not accurate for that setting any more. This also becomes apparent in the simulation study in Section 6.2 when the performance of the estimators is compared. The generalized multiscale estimator is unbiased if we disregard drift terms and particularly not biased due to asynchronicity. Furthermore, it passes over to the Hayashi-Yoshida estimator (3.2) for $M_N = 1$, a $\sqrt{N}$-consistent estimator in the complete absence of noise for that the stable central limit theorem 3.1 holds true. The key result of this section is that our estimation method achieves an improved convergence rate in the model with decreasing noise variances. It is obtained by a direct extension of the proof of Theorem 4.1 in Section 4.3 when replacing the moments of the noise processes. A similar extension for the one-scale estimator where we obtain the rate $N^{\frac{1}{6}+\frac{\alpha}{3}}$ for a subsample frequency $i_N = c_{sub}N^{\frac{2}{3}(1-\alpha)}$ holds

analogously.

Corollary 5.3.1. *Consider the model of Assumption 1, Assumption 2 and Assumption 3, but with noise variances* $\eta_X^2(N) = \zeta_X N^{-\alpha}$, $\eta_Y^2 = \zeta_Y N^{-\alpha}$, $0 < \alpha < 1$ *and constants* $0 < \zeta_X < \infty$, $0 < \zeta_Y < \infty$. *The generalized multiscale estimator* (4.2) *with* $M_N = c_{multi} N^{\frac{1}{2}-\frac{\alpha}{2}}$ *and optimal weights* (4.14) *converges stably in law to a mixed Gaussian limit:*

$$N^{\frac{1}{4}+\frac{\alpha}{4}} \left(\widehat{\langle X, Y \rangle}_T^{multi} - \langle X, Y \rangle_T \right) \overset{st}{\rightsquigarrow} \mathbf{N}\left(0, \mathbf{AVAR}^*_{multi}\right) \tag{5.12}$$

with the asymptotic variance

$$\begin{aligned}\mathbf{AVAR}^*_{multi} &= c_{multi}^{-3} \left(24 + 12\, \frac{I_X(T) + I_Y(T)}{T} \right) \zeta_X \zeta_Y + c_{multi}^{-1}\, \frac{12 \zeta_X \zeta_Y}{5} \\ &+ c_{multi} \frac{26}{35} T \int_0^T G'(t) (\sigma_t^X \sigma_t^Y)^2 (1 + \rho_t^2)\, dt \\ &+ c_{multi}^{-1} \frac{12}{5} \left(\zeta_Y \int_0^T (1 + I_Y'(t)) (\sigma_t^X)^2\, dt + \zeta_X \int_0^T (1 + I_X'(t)) (\sigma_t^Y)^2\, dt \right) .\end{aligned}$$

- **Dependent observation errors**
  A natural question that arises from Assumption 3 is what happens if the observation errors are not i. i. d. , but serially dependent. It turns out that the generalized multiscale estimator is still asymptotically unbiased and rate-optimal as long as some mixing conditions hold. However, the asymptotic variance is impacted by the dependence structure between the noise processes.
  Suppose that Assumptions 1, 2 and 3 are fulfilled, but instead of the i. i. d. condition we only impose that the noise processes are stationary and ϕ-mixing coefficients decay exponentially. Denote $c^X(l) = \mathbb{C}\mathrm{ov}\left(\epsilon_{t_0}^X, \epsilon_{t_l}^X\right)$, $0 \le l \le n$ and analogously $c^Y(l) = \mathbb{C}\mathrm{ov}\left(\epsilon_{\tau_0}^Y, \epsilon_{\tau_l}^Y\right)$, $0 \le l \le m$. The extension of the theory for the generalized multiscale estimator to stationary ϕ-mixing noise processes follows the same principles as has been carried out for the univariate MSRV estimator in Aït-Sahalia et al. [2009]. The main differences are that the additional asymptotically negligible bias does not appear at all in the setting where both noise processes are uncorrelated and that, even for stationary noise, we cannot insert values of the autocorrelation functions c^X, c^Y because of the general asynchronous sampling schemes.
  The overall estimation error is again split into the uncorrelated addends due to noise, discretization and cross terms that are asymptotically independent. The signal term only depending on the latent efficient processes is not affected by the serial dependence structure of the noise and Proposition 4.3.4 stays valid. The other parts of the asymptotic variance (4.8) change. For the error due to noise, that has been treated for i. i. d. noise in Proposition 4.3.1, we have

$$\mathbf{AVAR}_{\mathrm{n}}^{dep} =$$

$$
\begin{aligned}
p\text{-}\lim_{N\to\infty} \frac{M_N^3}{N} & \left(\sum_{i=1}^{M_N} \frac{\alpha_{i,M_N}^{opt}}{i} \sum_{j=1}^{M_N} \frac{\alpha_{j,M_N}^{opt}}{j} \left(\sum_{k=j}^{N} \sum_{r=i}^{N} \mathbb{C}\mathrm{ov}\left(\epsilon_{g_k}^X, \epsilon_{g_r}^X\right) \mathbb{C}\mathrm{ov}\left(\epsilon_{\lambda_{j-k+1}}^Y, \epsilon_{\lambda_{r-i+1}}^Y\right) \right) \right. \\
& \left. + \sum_{i=1}^{M_N} \frac{\alpha_{i,M_N}^{opt}}{i} \sum_{j=1}^{M_N} \frac{\alpha_{j,M_N}^{opt}}{j} \left(\sum_{k=j}^{N} \sum_{r=i}^{N} \mathbb{C}\mathrm{ov}\left(\epsilon_{\gamma_k}^Y, \epsilon_{\gamma_r}^Y\right) \mathbb{C}\mathrm{ov}\left(\epsilon_{l_{j-k+1}}^X, \epsilon_{l_{r-i+1}}^X\right) \right) \right) .
\end{aligned}
$$

In the synchronous bivariate setting the last inner double sums can be simplified to

$$
\sum_{q=-(n-j)}^{n-i} \left(c^X(q) c^Y(q+i-j) + c^X(q+i-j) c^Y(q) \right) .
$$

The second term due to noise, considered in Proposition 4.3.2 before, has here the stochastic limit

$$
\begin{aligned}
M_N \sum_{i=1}^{M_N} \frac{\alpha_{i,M_N}^{opt}}{i} \sum_{j=1}^{M_N} \frac{\alpha_{j,M_N}^{opt}}{j} & \left(\sum_{k=1}^{j-1} \sum_{r=1}^{i-1} \mathbb{E}\left[\epsilon_{g_k}^X \epsilon_{g_r}^X\right] \mathbb{E}\left[\epsilon_{\gamma_k}^Y \epsilon_{\gamma_r}^Y\right] \right. \\
& \left. + \sum_{k=N-j+1}^{N} \sum_{r=N-i+1}^{N} \mathbb{E}\left[\epsilon_{l_k}^X \epsilon_{l_r}^X\right] \mathbb{E}\left[\epsilon_{\lambda_k}^Y \epsilon_{\lambda_r}^Y\right] \right) .
\end{aligned}
$$

The errors due to cross terms times M_N

$$
\begin{aligned}
M_N \sum_{i=1}^{M_N} \frac{\alpha_{i,M_N}^{opt}}{i} \sum_{j=1}^{M_N} \frac{\alpha_{j,M_N}^{opt}}{j} & \left(\sum_{k=i}^{N} \sum_{r=j}^{N} \left(Y_{\gamma_k} - Y_{\lambda_{k-i+1}}\right) \left(Y_{\gamma_r} - Y_{\lambda_{r-j+1}}\right) \right. \\
& \left. \times \mathbb{C}\mathrm{ov}\left(\epsilon_{g_k}^X - \epsilon_{l_{k-i+1}}^X, \epsilon_{g_r}^X - \epsilon_{l_{r-j+1}}^X\right) \right) ,
\end{aligned}
$$

and the symmetric term, also converge in probability on the imposed assumptions, but a closed-form expression is only available for the synchronous case where the inner sums equal

$$
\begin{aligned}
\langle Y \rangle_T & \sum_{q=-j}^{i} \left(\min\left(q+i+1, j\right) - \max\left(0, q\right)\right) \\
& \times \left(c^X(q) + c^X(q-i+j) - c^X(q-i+1) - c^X(q+j+1) \right) + \mathcal{O}_p(1) .
\end{aligned}
$$

The asymptotic properties of the generalized multiscale estimator are the same as for an i. i. d. noise setting. The only difference is the appearance of the asymptotic variance that becomes more complicated involving autocorrelations to all possible lags, and in the non-synchronous case the particular lags in the single addends are unknown. This robustness to serial dependence of the noise is thanks to the fact that the multiscale estimator allocates most weight to lower frequencies where the dependence is extraneous. Different to the bias-corrected two scales estimator, the one-scale subsampling estimator (4.3) is also robust to dependent noise since only

one low subsample frequency is used. Therefore, an adjustment as provided for the TSRV estimator in Aït-Sahalia et al. [2009] is not necessary here.
That the estimation methods do not require uncorrelated observation errors but are robust to serial dependence is very important for applications to ultra high-frequency data. In Section 6.3 in Figure 6.12 the autocorrelation functions of trading data substantiate this importance.

- **Mutually correlated noise processes**
The part of Assumption 3 that both noise processes are mutually independent can be relaxed if one wants to allow for correlations $\mathbb{E}\left[\epsilon_{t_i}^X \epsilon_{\tau_j}^Y\right] = \eta_{X,Y}^{i,j}$ for t_i and τ_j located near each other, similarly as the dependence structure of each process separately considered before. In any case the generalized multiscale estimator is asymptotically unbiased and keeps its features. This does not necessarily hold true for the one-scale estimator that would have to be bias-corrected as the TSRV estimator in the univariate case. A particular interesting case could be to include synchronous observations where $\mathbb{E}\left[\epsilon_{t_i}^X \epsilon_{t_i}^Y\right] = \eta_{X,Y}$ and $\forall \tau_j \neq t_i : \mathbb{E}\left[\epsilon_{t_i}^X \epsilon_{\tau_j}^Y\right] = 0$. Then the one-scale estimator with subsample frequency i has a bias $\eta_{X,Y}((S^* + S^{**})/i)$ where S^* denotes the number of synchronous observations greater or equal to i and S^{**} smaller or equal $(N - i)$. Note that synchronous observation times have to appear as maxima and minima of sets constructed with the synchronization algorithm 3.1. Here, a bias-correction $((S^* + S^{**})/i) \cdot S^{-1} \sum \Delta X_{t_i} \Delta Y_{t_i}$ can be found explicitly when S denotes the number of all synchronous sampling times.

- **A model for latent efficient processes allowing for jumps**
From an application oriented perspective, most important issues for the estimation of integrated covariances from ultra high-frequency data are beyond doubt asynchronicity and market microstructure noise. However, it is well-known that continuous semimartingales cannot describe totally the dynamics of an underlying latent efficient log-price on a microscopic or macroscopic time-scale. For that reason, nowadays log-prices are most commonly modeled as general semimartingales including jumps. Jumps are caused especially by the inflow of market news. Applications of tests for jumps like the one developed in Aït-Sahalia and Jacod [2009] often reveal the presence of jumps for financial data. This has also been reported in Huang and Tauchen [2005], among others. That persuaded us to provide a possibility to allow for jumps in the efficient price model. Consider general semimartingales

$$X_t = \int_0^t \mu_s^X \, ds + \int_0^t \sigma_s^X dB_s^X + \sum_{l=1}^{J_t^X} L_l^X \ , \ Y_t = \int_0^t \mu_s^Y \, ds + \int_0^t \sigma_s^Y dB_s^Y + \sum_{l=1}^{J_t^Y} L_l^Y \ ,$$

with locally bounded drifts, continuous volatilities and counting processes J_t^X, J_t^Y counting the jumps of X and Y with jump sizes $L_l^X, l = 1, \dots, J_t^X$ and $L_l^Y, l = 1, \dots, J_t^Y$, respectively. We state without proof that the generalized multiscale estimator converges in probability to the total quadratic covariation

$$\int_0^T \rho_t \sigma_t^X \sigma_t^Y + \sum_{0 \le s \le t} \Delta X_s \Delta Y_s$$

where $\Delta X_s = X_s - X_{s,-}, \Delta Y_s = Y_s - Y_{s,-}$, and the second addend is the sum of the simultaneous co-jumps. This is in line with the findings for the univariate TSRV- and MSRV-estimators that converge to the quadratic variation $\int_0^t \left(\sigma_s^X\right)^2 ds + \sum_{l=1}^{J_t^X} \left(L_l^X\right)^2$ and analogously for $\tilde{Y}$ in Zhang et al. [2005], Zhang [2006], Fan and Wang [2007] and Christensen et al. [2010]. In that setting, however, one might be interested in disentangling the continuous part from the jumps. We propose to use a two-stage approach and adopt the strategy from Fan and Wang [2007] that provides a convincing extension to general semimartingales allowing for jumps in the univariate case. The procedure in the bivariate case is similar. First, using wavelet methods as presented in Fan and Wang [2007] that locate the jumps in the sample paths for both processes at times S_l^X, S_l^Y, jumps are detected and jump sizes estimated with

$$\hat{L}_l^X = \frac{\sum_{S_l^X \le t_i \le S_l^X + \gamma_n} \tilde{X}_{t_i}}{|S_l^X \le t_i \le S_l^X + \gamma_n|} - \frac{\sum_{S_l^X - \gamma_n \le t_i \le S_l^X} \tilde{X}_{t_i}}{|S_l^X - \gamma_n \le t_i \le S_l^X|} \ , \ l \in \{1, \dots, J_T^X\} \ ,$$

$$\hat{L}_l^Y = \frac{\sum_{S_l^Y \le t_i \le S_l^Y + \gamma_n} \tilde{X}_{t_i}}{|S_l^Y \le t_i \le S_l^Y + \gamma_n|} - \frac{\sum_{S_l^Y - \gamma_n \le t_i \le S_l^Y} \tilde{X}_{t_i}}{|S_l^Y - \gamma_n \le t_i \le S_l^Y|} \ , \ l \in \{1, \dots, J_T^Y\} \ .$$

The interval lengths γ_n^X, γ_n^Y will be chosen of order $\sqrt{n}$ as described in Fan and Wang [2007]. Filtering the jumps simply by subtracting the estimated jumps from $\tilde{X}$ and $\tilde{Y}$ leads to a consistent combined method when applying the original estimation approach to the jump-filtered values.

Except for the extension of the estimation to the general semimartingale case, where we have referred to the literature to find a combined method that allows for estimating the realized covariance without the impact of jumps, we have derived certain important model variations on the noise processes to that the generalized multiscale estimator is robust. The robustness of the method to different model specifications broadens its use in various applications. Other assumptions, as independence of noise and efficient processes, are crucial for the analysis throughout this work. A remark on possible further extensions of the model that might be interesting for future research, especially for applications in financial studies, is given in the Conclusion.

6 Simulation study and real data analysis

In this chapter the estimation methods that have been established in the Chapters 4 and 5 are examined in an application study.
In Section 6.2 the features of the estimators are analyzed in a simulation study especially focusing on finite sample size characteristics, tests on robustness to several model specifications and sensitivity on frequencies. Furthermore, it is shown in a comparison to the concurrent previous-tick kernel approach from Barndorff-Nielsen et al. [2008b] that the synchronization algorithm is particularly important when noise effects are present but of limited influence (cf. Section 5.3). For mild noise levels the resulting combined generalized multiscale estimator (4.2) outperforms methods that propose a synchronization procedure differing from the Hayashi-Yoshida approach.
Section 6.3 comes up with an application to real financial tick-data where we face several features of the data that can not be described accurately by a latent continuous semimartingale and microstructure noise. However, the estimation methods manage to reveal a certain systematic correlation structure between the financial time series and permit us to conclude about statistical inference.
First of all, we establish a feasible estimation procedure in Section 6.1 including rules to determine the tuning parameters.

6.1 Applying the estimation procedure: choice of tuning parameters

An implementation of the generalized multiscale estimator (4.2) and the one-scale subsample estimator (4.3) requires first a rule to choose the tuning parameters. For the histogram-based estimation of the asymptotic variances involving (5.2a)-(5.2d) all bin-wise multiscale frequencies and the number of bins have to be chosen. In this section a convenient and also not to complicated accurate algorithm to implement the estimators and obtain also estimates for the asymptotic variances is provided. From the theoretical considerations in Chapter 4, we have learned that a multiscale frequency $M_N = c_{multi}\sqrt{N}$ and a one-scale subsample frequency $i_N = c_{sub}N^{2/3}$ minimize the respective mean square errors. The constants appear in the addends of the asymptotic variances

$$\mathbf{AVAR}_{multi} = c_{multi}^{-3}\,\mathbf{AVAR}_{\mathrm{n}} + c_{multi}^{-1}\,\mathbf{AVAR}_{cross,\mathrm{n}} + c_{multi}\,\mathbf{AVAR}_{dis}\ ,$$

$$\mathbf{AVAR}_{sub} = c_{sub}^{-2}\,\mathbf{AVAR}_{\mathrm{n},sub} + c_{sub}\,\mathbf{AVAR}_{dis,sub}\ .$$

- Choose a priori L and calculate pilot estimator $\widehat{\mathbf{AVAR}}^{p}_{multi}$ with

$$\widehat{\mathbf{AVAR}}^{p}_{dis}=\frac{N}{2}\sum_{k=1}^{\lfloor N/L\rfloor}\left(\left(\tilde{X}_{g_{kL}}-\tilde{X}_{l_{(k-1)L+1}}\right)^2+\left(\tilde{X}_{g_{kL+2}}-\tilde{X}_{l_{(k-1)L+3}}\right)^2\right)\left(\tilde{Y}_{\gamma_{kL}}-\tilde{Y}_{\lambda_{(k-1)L+1}}\right)^2$$

and $\widehat{\langle X\rangle}^{p}_{T}=\sum_{j=kL,k\geq 1}\left(\tilde{X}_{t_j}-\tilde{X}_{t_{j-L}}\right)^2$ and $\widehat{\langle Y\rangle}^{p}_{T}$ analogously and $I_X(t)\equiv I_X^N(T)$ and $I_Y(t)\equiv I_Y^N(T)$. Calculate $\widehat{\eta^2_X}$ and $\widehat{\eta^2_Y}$ according to (5.1).

- Use pilot estimates to estimate optimal constant(s)

$$\hat{c}^{(p)}_{multi}=\left(\frac{-\widehat{\mathbf{AVAR}}^{p}_{cross,\mathrm{n}}+\sqrt{\left(\widehat{\mathbf{AVAR}}^{p}_{cross,\mathrm{n}}\right)^2+12\widehat{\mathbf{AVAR}}^{p}_{dis}\widehat{\mathbf{AVAR}}^{p}_{\mathrm{n}}}}{6\widehat{\mathbf{AVAR}}^{p}_{\mathrm{n}}}\right)^{-1/2} \tag{6.1}$$

and $\hat{c}^{(p)}_{sub}=\sqrt[3]{\dfrac{2\widehat{\mathbf{AVAR}}^{p}_{\mathrm{n},sub}}{\widehat{\mathbf{AVAR}}^{p}_{dis,sub}}}$.

- Calculate $\hat{I}_1-\hat{I}_4$, given in (5.2a)-(5.2d), with

$$K_N=\sqrt{\hat{c}^{(p)}_{multi}}N^{1/5}\ \text{ bins and }\ M_N(j)=\left(\hat{c}^{(p)}_{multi}\right)^{5/4}N^{3/5}\ \forall j\ .$$

- Estimate asymptotic variance with $\hat{I}_1-\hat{I}_4$ and $\widehat{\eta^2_X},\widehat{\eta^2_Y}$ and determine $\hat{c}_{multi}$ and $\hat{c}_{sub}$ with the above given formulae.
- Calculate the generalized multiscale estimator (4.2) with optimal weights (4.14) (and the one-scale subsample estimator) with $M_N=\hat{c}_{multi}\sqrt{N}$ (and $i_N=\hat{c}_{sub}N^{2/3}$)

Algorithm 6.1: Algorithm for the estimation procedure.

A plausible selection hereby can be derived as solutions of the minimization problems of the asymptotic variances although this choice is based on asymptotics and not the finite sample size distribution. The solution of the minimization for the generalized multiscale estimator is given in formula (6.1) in Algorithm 6.1. The resulting formula for the one-scale estimator is also stated in Algorithm 6.1. This tactic has been proposed in Zhang et al. [2005] for the original two scales realized volatility leading to an analogous solution as for the one-scale estimator here.
Since the terms in the asymptotic variances are random and unknown, we are in need of consistent estimators to apply these formulae. The idea is the following: If we had a priori an estimator for the discretization part and the asymptotic variance due to cross terms and end-effects, together with the estimators for the noise variances presented at the beginning of Chapter 5, we obtain a pilot estimate $\hat{c}^{(p)}_{multi}$ for c_{multi} as solution of

formula (6.1). With this estimate we set up the estimation of the asymptotic variance involving the estimators (5.2a)-(5.2d). We proceed and set $M_N^b = c^b_{multi}\sqrt{NK_N}$ fixed for the multiscale estimators on all bins. We take the optimal order of bins K_N for common stochastic volatility models from Section 5.1 so that $M_N^b = c^b_{multi}\sqrt{Nc_K N^{1/5}}$. Now we use $\hat{c}^{(p)}_{multi}$ and set $c^b_{multi}c_K^{-1/2} = \hat{c}^{(p)}_{multi}$ and from the orders of the different errors of the histogram estimators $c^b_{multi} = c_K^{5/2}$. Hence, $c^b_{multi} = \left(c^{(p)}_{multi}\right)^{5/4}$ and $c_K = \sqrt{c^{(p)}_{multi}}$ is derived. We use the same tuning parameters for all estimators (5.2a)-(5.2d) though there might possibly be differences between the (co-)variations of times and the degrees of regularity of asynchronicity. Anyway, this will neither be the case in typical applications nor have the estimators (5.2c) and (5.2d) a deep impact on the total estimates for the asymptotic variance. Using estimators (5.2a)-(5.2d), we calculate estimates for the addends of the asymptotic variance and $\hat{c}_{multi}$ according to formula (6.1) again. The multiscale frequency $M_N = \lceil \hat{c}_{multi}\sqrt{N}\rceil$ is used to evaluate the final estimator for the quadratic (co-)variation. We proceed analogously for the univariate case and the one-scale estimator. It remains to choose pilot estimates for the quadratic variations and the discretization part of the asymptotic variance. For this purpose we take sparse-sampled versions of the estimators that are consistent in the absence of microstructure noise. The only parameter that still has to be set is the sparse-sample frequency that can be chosen under the impression of the signature plots, but a choice of integers in a wide domain is appropriate to initialize the procedure effectively. Even though the procedure might seem somewhat arbitrary and is not unique, it turns out that it is very robust to the a priori sparse-sample frequency and performs well in the following application study.

6.2 Simulation study

In this section the finite sample size characteristics of the estimation approach that has been proposed in Chapter 4 are investigated. The main features of the one-scale subsampling (4.3) and the generalized multiscale estimator (4.2) are analyzed in a parametric model and the estimation procedure given in Algorithm 6.1 is applied. As a benchmark for comparison, the kernel estimator as presented in Barndorff-Nielsen et al. [2008b] is considered, too. The multiscale estimator is further tested for its robustness to more realistic models. Parts of this simulation study have been published in Bibinger [2011]. Asynchronous observation times are generated as arrival times of two mutually independent homogeneous Poisson processes on $[0, 1]$.
For the beginning a Brownian motion model with constant parameters $\sigma^X = \sigma^Y = 1$, $\rho \in [-1, 1]$ and zero drifts is implemented. The increments of the efficient processes

$$\Delta X_{t_i} = \int_{t_{i-1}}^{t_i} dB_t^X \;, i \in \{0, \dots, n\}\;,$$

and

$$\Delta Y_{\tau_j} = \int_{\tau_{j-1}}^{\tau_j} dB_t^Y = \int_{\tau_{j-1}}^{\tau_j} \rho dB_t^X + \sqrt{1-\rho^2}\int_{\tau_{j-1}}^{\tau_j} dB_t \;, j \in \{0, \dots, m\}\;,$$

$\mathbb{E}[n]$	$\mathbb{E}[m]$	sd	$\sqrt{\mathbf{AVAR}_{HY}}$
23400	23400	0.0143	0.0142
23400	11700	0.0173	0.0175
23400	5850	0.0229	0.0227
23400	2925	0.0311	0.0307

Table 6.1: Standard deviations of Hayashi-Yoshida estimates and comparison to calculated asymptotic values.

where B_t is a standard Brownian motion independent of X are simulated on the grid set by the Poisson schemes.

First, we compare the finite sample variances of the Hayashi-Yoshida estimator in this parametric model to the values of the asymptotic variances deduced in Sections 3.2, 3.3 and 5.2. Table 6.1 gives the empirical standard deviations (sd) of 1000 Monte Carlo iterations and the asymptotic limits calculated according to the results of Section 5.2 for different values $\mathbb{E}[n]$ and $\mathbb{E}[m]$ of the expected numbers of observations of X and Y. Simulations of the Hayashi-Yoshida estimator under noise can be found in Palandri [2006], among others. From now on set the expectations of the time increments between two observations for X and Y equally to 1/30000 and $T = 1$.

For an illustration of the gain of the synchronization method according to Algorithm 3.1 and the estimator (3.2) compared to a previous-tick interpolated realized covariance, we have implemented both for this parametric model and Poisson sampling. When $\rho = 0.5$, the estimators based on 200 Monte Carlo iterations have means 0.501 and 0.335, respectively, with standard deviations 0.0128 and 0.0088. The negative bias aroused by the previous-tick steps, that has an expectation 1/6, is obvious.

Next, we add Gaussian noise $\epsilon_{t_i}^X \sim \mathcal{N}(0, \eta_X^2)$ and $\epsilon_{\tau_j}^Y \sim \mathcal{N}(0, \eta_Y^2)$ to the Brownian motions. The resulting root mean square errors (RMSE) based on 1000 Monte Carlo iterations calculated for different noise level are illustrated in Figure 6.4. The one-scale subsample and the multiscale estimator have been evaluated with the subsample frequencies (SSFR) and multiscale frequencies (MSFR) listed in Table 6.2. These are calculated with the formulas in Algorithm 6.1 and the feasible parameter values. The estimators (5.2a)-(5.2d) and the resulting estimates for the asymptotic variances and values of the one-scale and the multiscale estimator after an application of the complete adaptive estimation procedure according to 6.1 are given in Table 6.2.

The previous-tick refresh time kernel estimator is simulated with the Parzen kernel and 'jittering' to handle end-effects. This means that certain numbers of values at the beginning and at the end are averaged, where we have chosen that number after heuristic optimization for each noise level taking values between 4 and 8. The optimal bandwidth can be calculated for our simple model. For more information on the kernel method we refer to Barndorff-Nielsen et al. [2008a] and Barndorff-Nielsen et al. [2008b].

In Chapter 4 it has been proved that the rate of convergence $N^{1/4}$ of the multiscale estimator is much faster compared to $N^{1/6}$ for the one-scale estimator. However, simulation experiences in the literature indicated that the differences of finite sample variances are rather small in the univariate case. Figure 6.4 exemplifies that for very high noise levels

Figure 6.2: Calculated frequencies for different noise levels $\eta_X^2 = \eta_Y^2$ and parameters $\sigma^X = \sigma^Y = 1, \rho = 1/2$.

noise level $\eta_X^2 = \eta_Y^2$	SSFR	MSFR
$(1/\sqrt{10})$	627	289
0.1	291	164
$(1/\sqrt{10})\cdot$ 0.1	135	93
0.01	63	52
$(1/\sqrt{10})\cdot$ 0.01	29	29
0.001	17	14
$(1/\sqrt{10})\cdot$ 0.001	9	6
0.0001	5	3

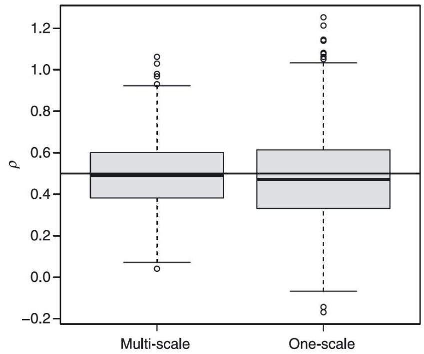


Figure 6.3: Boxplot of 1000 realizations for the generalized multiscale and the one-scale subsampling estimator for $\rho = 0.5$ when $\eta_X^2 = \eta_Y^2\sqrt{0.1}$.

and $\mathbb{E}N = 20000$, the generalized multiscale estimator has a significant smaller root mean square error than the one-scale estimator. When the noise level decreases the differences vanish. The kernel estimator exhibits another effect resulting in an increasing root mean square error for small noise variances. This is caused by the synchronization technique of the method and the bias due to asynchronicity. The bandwidths calculated for these noise levels take small values and so the effect increases. This confirms that for limited noise contamination one can only end up with an improved estimation when using the synchronization from Algorithm 3.1.

The deviation of the estimated asymptotic variances and the calculated values in Table 6.2, some being off the standard empirical confidence sets, is partly explained by the fact that we included optimal subsample and multiscale frequencies in the calculation whereas the estimates use adaptive choices. After all, the methods provide satisfying estimates and allow for statistical inference. The finite sample size variances are slightly greater than the asymptotic ones. The noise levels in Figure 6.4 are chosen such that the asymptotic variance due to noise terms is linearly increasing. This can be seen for larger noise variances when the error due to noise dominates the signal term.

In Figure 6.5 the root mean square errors of the three estimators are diagrammed for different constant parameter values of the correlation $\rho = k/10\,, k = 0, \ldots, 10$ and a fixed noise variance $\eta^2 = 0.01$ based on 200 Monte Carlo iterations. The increasing root mean square errors when ρ increases are explained by the variance parts due to discretization that are similar for all three estimators. At this noise level, the discretization error is influential enough to cause the increasing root mean square errors illustrated in Figure 6.5.

The simple parametric model which we have implemented here enables us to reveal the finite sample characteristics of the estimators. Further simulations provided below attest that the generalized multiscale estimator performs well also in more realistic models.

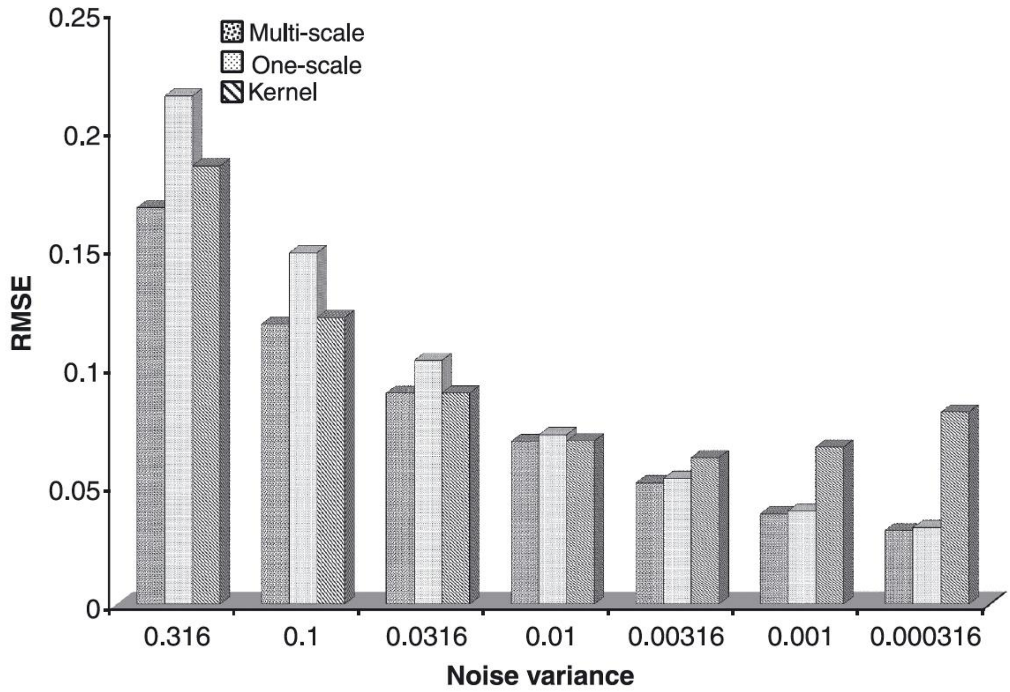


Figure 6.4: Root mean square errors of the one-scale, the generalized multiscale and the kernel estimator for different noise levels $\eta_X^2 = \eta_Y^2 = \eta^2$ for $\rho = 0.5$.

Sensitivity to choice of frequencies

Next, we take a look at implementations of the one-scale and the multiscale estimator for varying frequencies to analyze the dependence on the associated subsampling and multiscale frequencies, or rather the constants $c_{sub} = i_N \cdot N^{-2/3}$ and $c_{multi} = M_N/\sqrt{N}$. Since we use an adaptive selection rule based on minimization of the asymptotic variances and the estimators (5.2a)-(5.2d), it is important to learn about the robustness to the frequencies and how the estimators might react to misspecified choices of the constants. For this purpose, we run a simulation study of the Brownian model as before with constant correlation $\rho = 1/2$ and additive Gaussian i. i. d. noise. We fix the noise level $\eta^2 = 1/\sqrt{10} \cdot 0.1$. From 100 Monte Carlo iterations, Figure 6.6 illustrates the root mean square errors of the one-scale and the multiscale estimator for a large domain of frequencies $i_N = M_N = l \cdot 50\,,\, l \in \{1, \ldots, 10\}$. Table 6.2 gives the values $M_N = 93$ and $i_N = 135$ according to our selection rule based on asymptotics and inserting the known parameters. The main finding is that both estimators are not too sensitive to the choice of the frequency or according constants. Only if those tuning parameters are chosen too small the root mean square errors increase. The frequency choice based on asymptotics proposed in Algorithm 6.1 seems to be adequate but in case of doubt should be rather chosen bigger and not smaller what is in accordance with our experience.

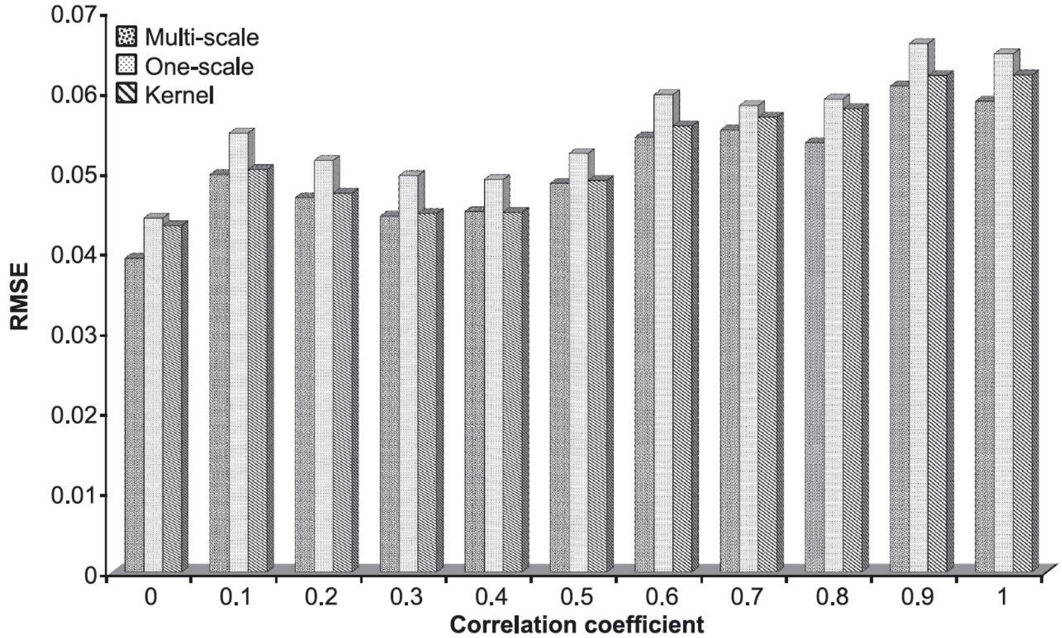


Figure 6.5: Root mean square errors of the one-scale, the generalized multiscale and the kernel estimator for a constant noise level $\eta^2 = 0.01$ and different correlation coefficients.

Robustness to non–i. i. d. noise

In this paragraph the robustness of the generalized multiscale estimator to serial dependence in the noise processes is examined. In Section 5.3 we have sketched the impact on the asymptotic variance and claimed that consistency and rate-optimality of our proposed estimator stay valid when relaxing the i. i. d. condition to a less restrictive and more realistic assumption of stationary discrete processes with exponentially decaying ϕ-mixing coefficients. To accentuate this property, we have implemented an Itô diffusion model with a constant correlation coefficient $\rho = 1/2$, $T = 1$, constant spot volatilities $\sigma^X = \sigma^Y = 1$ and $\mu^X = \mu^Y = 0$, but where the microstructure noise is following a martingale difference model. In particular, we generate the observation errors according to

$$\epsilon^X_{t_i} \sim \mathbf{N}\left(-\alpha\epsilon^X_{t_{i-1}}\,,\,(\eta^X + 1/30000 - \beta\Delta X^2_{t_{i-1}})\right)\ ,$$

$$\epsilon^Y_{\tau_j} \sim \mathbf{N}\left(-\alpha\epsilon^Y_{\tau_{j-1}}\,,\,(\eta^Y + 1/30000 - \beta\Delta Y^2_{\tau_{j-1}})\right)\ .$$

This accomplishes a negative correlation of succeeding observation errors and a dependence of the noise variances on the evolution of the efficient processes in the preceding observation time instant. It therefore has some important realistic market features and a satisfying performance in this model is desirable.
In Figure 6.8 the root mean square errors from 100 iterations of the generalized multiscale estimator are illustrated for two different noise levels and different values of α for constant $\beta = 100$. Table 6.3 contains the bias and standard deviations for a higher $\beta = 1000$ and the higher noise level. The root mean square errors are increasing with α, but even for large values of α, e. g. $\alpha = 6/7$, the performance of the generalized multiscale estimator is still satisfying. It is thus confirmed that the estimator is quite robust to serially dependent noise.

A stochastic volatility model

In this paragraph we apply our one-scale and the generalized multiscale estimator to a stochastic volatility model. We have implemented a similar model as Barndorff-Nielsen et al. [2008b] and Huang and Tauchen [2005] before. Mutually independent homogeneous Poisson sampling schemes with same intensities are incorporated as before. X and Y are defined by

$$dX = \mu^X dt + \rho^X \exp\left(\beta_0^X + \beta_1^X \varrho^X\right) dB^X + \sqrt{1-(\rho^X)^2}\exp\left(\beta_0^X + \beta_1^X \varrho^X\right) dB^{\perp} \ ,$$

$$dY = \mu^Y dt + \rho^Y \exp\left(\beta_0^Y + \beta_1^Y \varrho^Y\right) dB^Y + \sqrt{1-(\rho^Y)^2}\exp\left(\beta_0^Y + \beta_1^Y \varrho^Y\right) dB^{\perp} \ .$$

with the Ornstein–Uhlenbeck process

$$d\varrho^X = \alpha^X \varrho^X dt + dB^X \ .$$

Define the process ϱ^Y analogously. There is a leverage given by ρ^X and ρ^Y, respectively. The correlation coefficient $\sqrt{1-(\rho^X)^2}\sqrt{1-(\rho^Y)^2}$ is constant and $B^{\perp}$ is a Brownian motion, independent of B^X and B^Y, driving the common factor. The discrete observations are simulated via an Euler scheme. For the noise we take Gaussian i. i. d. errors as above. The constants are chosen exactly as in Barndorff-Nielsen et al. [2008b] to guarantee a good comparability of the results. The values are set as follows:

$$\mu^X = \mu^Y = 0.03;\ \alpha^X = \alpha^Y = -1/40;\ \beta_0^X = \beta_0^Y = -5/16;\ \beta_1^X = \beta_1^Y = 1/8;$$
$$\rho^X = \rho^Y = -0.3 \ .$$

We use an exact discretization of the Ornstein–Uhlenbeck process on a subgrid with instants $\delta = 1/300000$ on the interval $[0, 1]$. In Table 6.4 the mean absolute errors and the mean square errors for both estimators and three different noise levels are given based on 100 iterations. For a very small noise variance $\eta^2 = 0.001$, the error of the simpler one-scale estimator is smaller than for our generalized multiscale estimator, but when the influence of the noise increases the generalized multiscale approach outperforms the one-scale estimator as can be seen in Table 6.4 for the two larger noise levels.

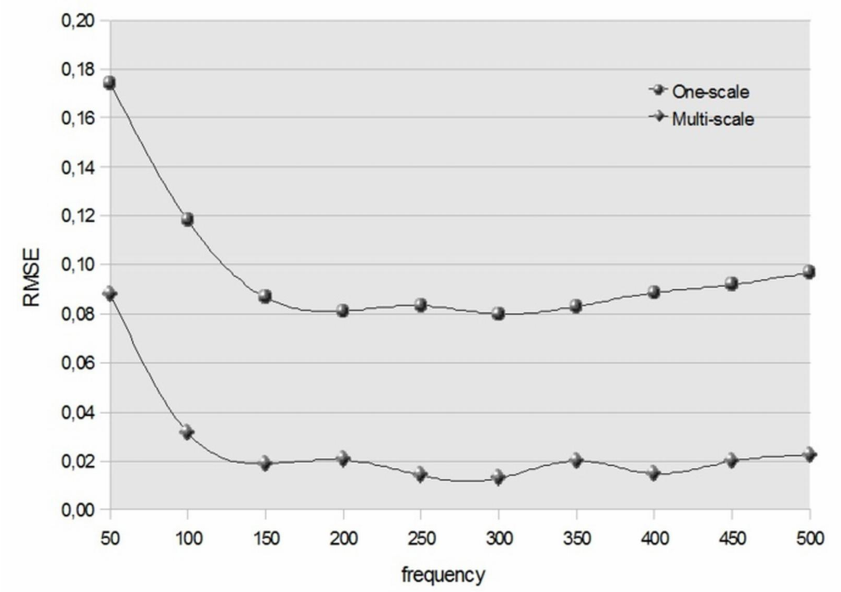


Figure 6.6: Sensitivity of the one-scale and the multiscale estimator to its tuning parameters or frequencies for a fixed noise variance $\eta^2 = 0.1/\sqrt{10}$.

6.3 Application to EUREX future tick-data

6.3.1 Data description

In the last section, we have analyzed the finite sample size behaviour of our estimation approach for simulated data according to the additive noise model that we have considered in Chapter 4.

The second part of the application study entails testing the methodology in a real data analysis. For this purpose we use an EUREX trading-database provided by the Research Data Center (RDC) of the CRC 649 'Economic Risk' in Berlin. The database contains all tick-data up to May 2008.

We apply the generalized multiscale estimator (4.2), including the data-aggregation with Algorithm 3.1 after the procedure given in Algorithm 6.1 to determine the tuning parameter, to estimate integrated covariances between the four financial securities with the highest tick-frequencies in the database. These are the Euro-Bund Future (FGBL), that is based on a notional long-duration debt instrument issued by the Federal Republic of Germany, the Euro-Bobl Future (FGBM), a likewise medium-duration contract, and futures on the EURO STOXX 50 (FESX) and the German DAX (FDAX). For each of them there are four expire dates every year and we sample out the one at hand for the considered days. We choose two ordinary days arbitrarily, January 10th and April 23rd

Table 6.2: Estimators (5.2a)-(5.2d), estimators for the asymptotic variances of the multiscale (4.8) and the one-scale estimator (4.10), calculated asymptotic variances and estimates for the quadratic covariation. The estimates are given ± empirical standard deviations.

noise var. η^2	0.0001	$(0.001/\sqrt{10})$	0.001	$(0.01/\sqrt{10})$	0.01
$\hat{I}_1$	0.392 ± 0.038	0.390 ± 0.047	0.394 ± 0.073	0.413 ± 0.144	0.423 ± 0.128
$\hat{I}_2$	1.557 ± 0.067	1.552 ± 0.085	1.538 ± 0.141	1.529 ± 0.276	1.462 ± 0.230
$\hat{I}_3$	0.250 ± 0.007	0.249 ± 0.010	0.249 ± 0.016	0.247 ± 0.031	0.234 ± 0.085
$\hat{I}_4$	0.250 ± 0.007	0.249 ± 0.009	0.249 ± 0.016	0.247 ± 0.030	0.233 ± 0.083
$\widehat{\mathbf{AVAR}}_{multi}$	0.090 ± 0.003	0.143 ± 0.005	0.246 ± 0.015	0.434 ± 0.050	0.778 ± 0.082
$\mathbf{AVAR}_{multi}$	0.0663	0.1185	0.2159	0.3774	0.6737
$\widehat{\mathbf{AVAR}}_{sub}$	0.0086 ± 0.0002	0.017 ± 0.001	0.037 ± 0.002	0.080 ± 0.009	0.157 ± 0.017
$\mathbf{AVAR}_{sub}$	0.0077	0.0166	0.0357	0.0768	0.1656
$\widehat{\langle X,Y\rangle}_T^{multi}$	0.501 ± 0.024	0.499 ± 0.029	0.498 ± 0.038	0.499 ± 0.049	0.501 ± 0.065
$\widehat{\langle X,Y\rangle}_T^{sub}$	0.500 ± 0.022	0.499 ± 0.028	0.499 ± 0.042	0.500 ± 0.058	0.503 ± 0.074

Table 6.3: Bias and standard deviation of the generalized multiscale estimator in the correlated noise model for different noise levels and values of α and β, respectively.

$\eta_X^2 = \eta_Y^2$	α	β	BIAS_{multi}	SD_{multi}
0.001	1/4	100	-0.00195	0.03640
·	1/3	100	0.00158	0.03309
·	1/2	100	-0.00471	0.03702
·	2/3	100	0.00085	0.04982
·	3/4	100	0.00634	0.05151
·	6/7	100	-0.01495	0.08456
$0.1/\sqrt{10}$	1/4	100	0.00717	0.05693
·	1/3	100	-0.00420	0.06074
·	1/2	100	-0.00952	0.07252
·	2/3	100	0.01058	0.08883
·	3/4	100	-0.00273	0.09265
·	6/7	100	0.01629	0.17034
$0.1/\sqrt{10}$	1/4	1000	0.00022	0.07445
·	1/3	1000	-0.01792	0.08190
·	1/2	1000	-0.00442	0.08349
·	2/3	1000	-0.00572	0.08802
·	3/4	1000	-0.01621	0.12891
·	6/7	1000	-0.02856	0.16866

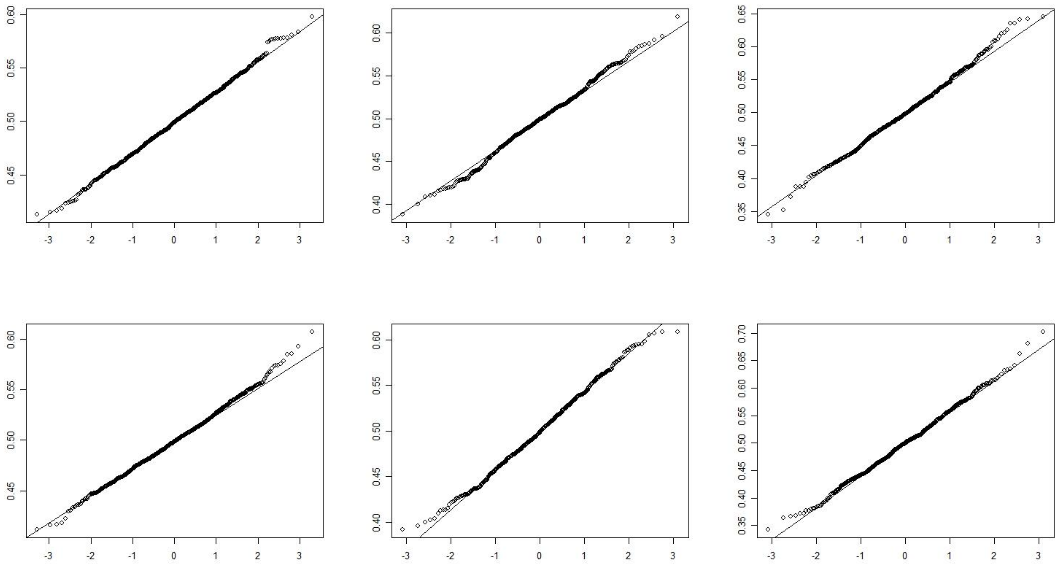


Figure 6.7: Normal Q-Q plots of estimates from 1000 Monte Carlo iterations for the quadratic covariation: multiscale (top-line), one-scale (bottom-line), noise levels $\eta = 0.01778, 0.03162, 0.05623$ in first, second, third row.

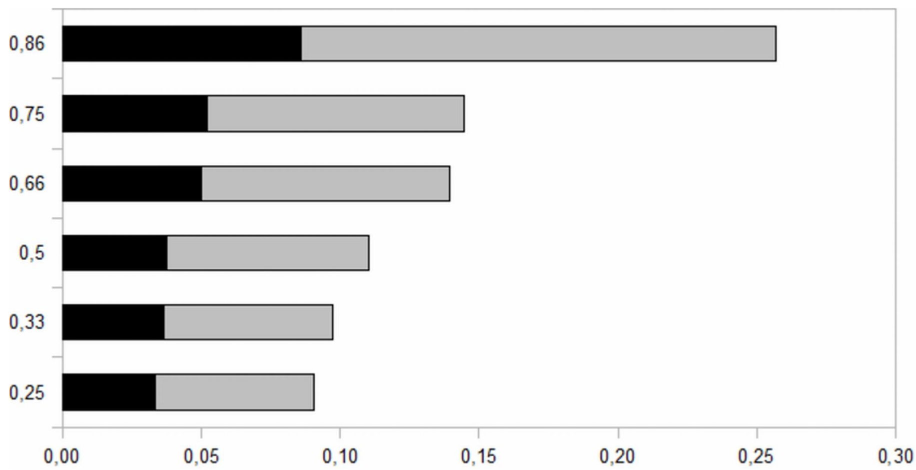


Figure 6.8: Root mean square errors for correlated noise model for different values of α (y-axis) and two noise levels.

in 2008, and as third sample day September 11th, 2001. As time span we set 8 am – 4 pm for the first two days and 8 am – c. 5.30 pm, when the last tick of the ESX was recorded, for 09/11 (all times in CET).
The documentation schemes of trading events (recorded on a 0.00001 second-grid) list also

Table 6.4: Simulation results for the one-scale and the multiscale estimator in the stochastic volatility model.

Noise variance	mean abs. error$_{multi}$	mean abs. error$_{one}$	MSE$_{multi}$	MSE$_{one}$
0.001	0.07001	0.05370	0.00763	0.00491
$0.1/\sqrt{10}$	0.08373	0.11024	0.01042	0.01722
0.1	0.09487	0.11504	0.01135	0.01891

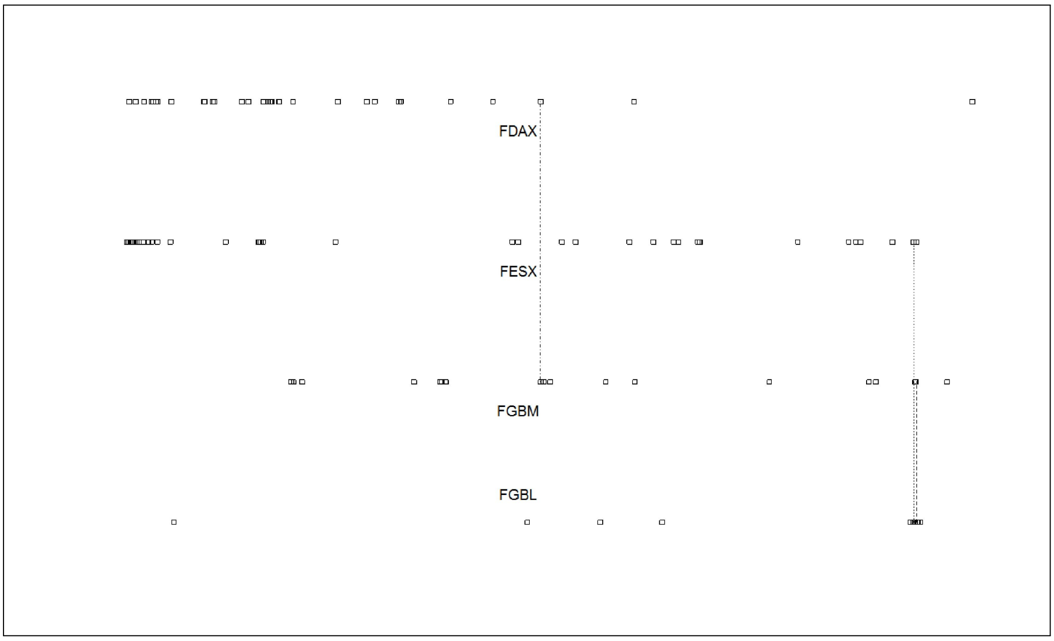


Figure 6.9: Example for observation schemes, 04/23/2008, 15 seconds interval starting at 3 pm.

synchronous observations between the assets, but most observations are non-synchronous. We exemplify the observation schemes in Figure 6.9 where synchronous ticks are marked with dashed connectors. Since we have used a characteristic signature plot from that data sample (FESX) to motivate the economic background of the additive noise model in Figure 0.1 in the Introduction and the autocorrelations for the differences of log-prices (returns) in Figure 6.12 show a typical MA(1) appearance as in a Brownian i. i. d. noise model, main features of the statistical model considered in Chapter 4 should fit to the data and our methods provide a convenient way to estimate integrated covariances. However, there are several limitations where the characteristics of the data harm the accordance with the model assumptions:

1. Price discreteness (cent for FGBL/FGBM; 0.5€ for FESX; € for FDAX).

2. Most returns are zero.

3. Sample paths not injective (different prices recorded at same time).

4. Trading occurs at discrete times since observations take place on discrete grid (see above).

5. Limited sample sizes, about 13000 ticks for the lowest frequent FGBM and c. 40000 ticks for the highest frequent FDAX, and only c. 3000 - c. 20000 non-zero returns (transaction data).

6. Certain 'outliers' appear as can be seen in Figure 6.11 that could be due to documentation problems (delays etc.) and specific trading conditions and principles.

Especially the first and second point are inconsistent with the imposed model of an underlying efficient log-price that is described by a continuous semimartingale with additive noise.
The quadratic (co-)variations of time and the degrees of regularity of asynchronicity are plotted for 04/23 and FGBL/FGBM as well as 09/11 and FESX/FDAX in Figure 6.10. These graphics reflect the fact that observation times do vary more than e. g. in the Poisson case on the one hand and that on the other hand the difference quotients appearing in the variances not necessarily tend to a globally constant limit. The longest intervals where no trading has been recorded in the considered data are about one minute whereas certain periods of the day feature with several observations per second in average. An increasing trading frequency arises from the opening of the US markets after lunch time and affects the plotted functions in Figure 6.10. Since the ESX and the

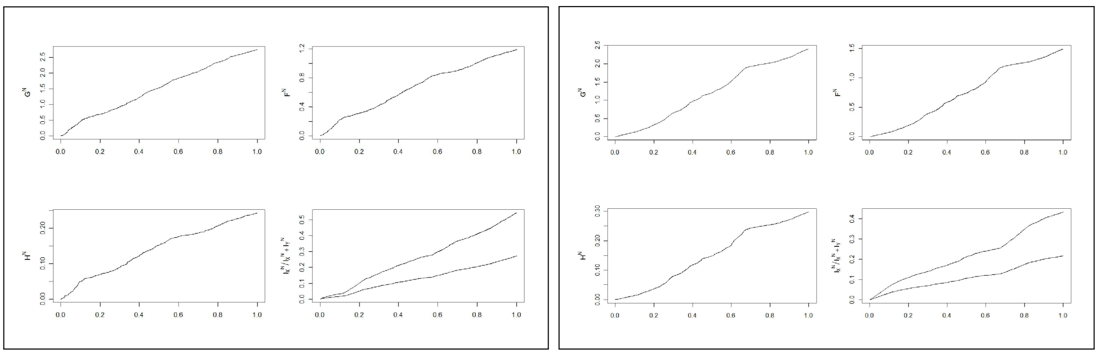


Figure 6.10: Quadratic (co-)variations of times for FGBL/FGBM, 04/23/2008, and FESX/FDAX, 09/11/2001.

DAX share 13 companies constituting c. 28.5% weighting in the ESX and c. 72.4% in the DAX there is a big systematic positve correlation between both. From Figure 6.11 and the close relation of FGBM and FGBL, we can presume that there is as well a high correlation between those. On 09/11 we see a tremendous effect in the afternoon that FGBL/FGBM increase and the FESX/FDAX decrease. The range of FESX and FDAX for 09/11 in Figure 6.11 is about five times the one for the day illustrated above.

6.3.2 Estimation procedure

We apply the generalized multiscale estimator (4.2) with a multiscale frequency $M_N = \widehat{c}_{multi}\sqrt{N}$ determined by Algorithm 6.1 and, hence, obtain also estimates for the asymptotic variance. Integrated volatilities are estimated with MSRV estimators following an analogous selection rule for the used multiscale frequencies. The 'outliers' mentioned above are kept in the data samples. Note that unless they are only very few, these returns slightly influence the estimators, especially the univariate ones and the simple realized volatilities in the signature plot (Figure 0.1). For the bivariate estimators there is a difference according to whether one uses tick or transaction data. We keep to the tick-data, but we adjust the estimators for the noise variances (5.1) to $(2n^*)^{-1}\mathbf{RV}$ where $\mathbf{RV}$ is the realized volatility at the highest available frequency and n^* the number of non-zero returns. It turns out that the estimated noise variances still will be rather small compared to the estimated discretization variances. In the solution of the minimization of the asymptotic variance leading to (6.1), the variance due to cross terms and end-effects has the smallest influence on $\widehat{c}_{multi}$ and has hardly effected the estimates in the simulation part. For the data analysis, however, when estimated noise variances are quite small the estimate for that addend of the variance can sometimes dominate the product of the two others in the discriminant in formula (6.1) and pushes the constant to be chosen too small what results in a multiscale frequency equal to one (Hayashi–Yoshida estimator). For that reason it is convincing to remove that term and choose the tuning parameter only according to the balance of noise and discretization variance.

In summary, we use exactly the approach as stated in Algorithm 6.1 with two small adjustments that $\widehat{\mathbf{AVAR}}_{cross,n} = 0$ and we estimate noise variances with the number of non-zero ticks.

6.3.3 Results

The pilot estimates of the multiscale frequency are already very robust to the a priori chosen sparse-sampling frequency since the estimators of the noise variances are dominant. In Table 6.6, the resulting estimates for the asymptotic variances for all combinations and the three chosen days are listed together with the number of bins and the fixed multiscale frequency for the binwise evaluated estimators (5.2a) and (5.2b) in parenthesis. Our selection rule gets us to choose the number of bins for all cases equal to one or equal to two. Actually, the estimates are also very robust to different values K_N of bins and multiscale frequencies M_N for the estimators evaluated on each bin. For FGBL and FGBM for example on 23rd April 2008 all different choices $1 \leq K_N \leq 10$ with fixed $M_N = 5$ as in Table 6.6 and as well for larger $M_N = 50$ lead to almost the same estimates, in every case $6.399 \cdot 10^{-11}$ rounded to the last given figure.

The estimated integrated covariances and the integrated volatilities are given in Table 6.5 together with the chosen multiscale frequency (MSFR) by Algorithm 6.1. Chosen between three and fifteen they are rather small compared to the values experienced in the simulation study. Since the frequencies are chosen smaller for small noise levels those values are not that surprising for the considered data. For the univariate integrated volatilities the MSRV (2.11b) and also the TSRV estimators (2.9b) are robust against choosing different frequencies in a wide domain. In Figure 6.13 we illustrate the TSRV estimators calculated with different subsampling frequencies for the 23rd April 2008. In an asynchronous setting the generalized multiscale estimator and the one-scale subsampling estimator, plotted in Figure 6.14 against their sampling frequencies, are not that robust that the selection of the tuning parameter does not have an effect. The curves in Figure 6.14 for the FGBL/FGBM tick-data, 04/23, FGBL/FGBM transaction data for the same day and FGBL/FGBM and FESX/FDAX tick-data, 01/10, look very similar to the ones we obtain for data simulated from the additive noise model. For low frequencies the variance due to noise explodes and for too large frequencies the discretization error dominates. Both tend downwards or upwards, but are correlated for different frequencies what explains why we do not see an oscillation. The effect seen for low frequencies in the first graphic of Figure 6.14 has nothing to do with an Epps effect, all estimators are adapted in the fashion of the Hayashi-Yoshida estimator to asynchronicity. The effect of decreasing estimates for low frequencies vanishes for the transaction data in Figure 6.14. If one would suggest to choose the multiscale frequencies in the area where the first curve has its maximum, this rule cannot be maintained when the curve is monotone as in the third graphic. The comparison with the transaction data shows, furthermore, that this maximum of the tick-data is not present for transaction data. As can be seen in Table 6.5 according to our method we select $M_N = 6$ and an estimate $4.81 \cdot 10^{-6}$ for 04/23 and $M_N = 7$ and, hence an estimate $3.32 \cdot 10^{-6}$ for 01/10.

From the stable asymptotic mixed normality result in Theorem 4.1, we can deduce that the estimation error divided by the estimated standard deviation weakly converges to a standard normal distribution. This allows directly to give an asymptotic distribution free test for the hypothesis that $\langle X, Y \rangle_T = 0$. The test statistics and the corresponding p-values for a two-sided test are listed in Table 6.7. Besides the integrated volatilities, we can also reject the null hypothesis for the integrated covariances of the FGBL/FGBM and the FESX/FDAX for all three days to any reasonable size of test, what one might have expected.

Between the dept bonds and the index futures on 04/23 and 01/10 all except one integrated covariances are estimated greater than zero, but the p-values to reject the null are at least 0.2. For 09/11 things look different and we get, as presumed, negative estimated integrated covariances and for the FGBL/FESX and FGBL/FDAX p-values less than 5% and for FGBM/FESX c. 6.9%.

6.4 Conclusion of the application study

In conclusion, the methods with very little convenient adjustments perform well in a real data analysis although there are some obvious limitations where the considered model cannot fit the data very accurately. The nature of the chosen securities allowed to test if the methods help to quantify effects that could be foreseen. The generalized multiscale estimator depends on a tuning parameter that has to be selected first what might be a

drawback compared to methods that get along without as Aït-Sahalia et al. [2010].
Anyway, Algorithm 6.1 gives a straight selection rule that is reliable for data applications where the considered data is not too far away from the additive noise model. For the moment, the proposed method is the only one allowing for statistical inference by the feasible stable central limit theorem and bridges the gap to the non-noisy setting. We emphasize that no bias due to interpolations and asynchronicity occurs.

In the simulation study in Section 6.2 it has been shown that the generalized multiscale estimator outperforms the kernel-approach with refresh times for mild noise levels as typically present in financial applications. In the case of higher noise corruption both methods perform comparably well. The study has also confirmed that the generalized multiscale estimator is robust to several varieties of the model, also for finite sample sizes.

	$\widehat{\langle X,Y\rangle}_T^{multi}$	FGBL	FGBM	FESX	FDAX
04/23/2008	FGBL	8.95 (9)	4.81 (6)	1.31 (10)	0.45 (6)
	FGBM		3.92 (5)	-0.07 (8)	0.51 (4)
	FESX			89.70 (10)	24.95 (7)
	FDAX				74.42 (4)
	$\widehat{\langle X,Y\rangle}_T^{multi}$	FGBL	FGBM	FESX	FDAX
01/10/2008	FGBL	5.46 (7)	3.32 (7)	0.98 (9)	0.52 (6)
	FGBM		2.42 (6)	0.78 (7)	0.64 (4)
	FESX			68.26 (10)	29.39 (7)
	FDAX				61.43 (5)
	$\widehat{\langle X,Y\rangle}_T^{multi}$	FGBL	FGBM	FESX	FDAX
09/11/2001	FGBL	27.89 (15)	12.94 (12)	-52.55 (8)	-34.13 (7)
	FGBM		18.10 (8)	-26.44 (6)	-25.01 (3)
	FESX			3070 (6)	757 (4)
	FDAX				1870 (4)

Table 6.5: Estimates for integrated covariances ($\cdot\ 10^6$) and used multiscale frequencies with used multiscale frequency (MSFR).

Date	$\widehat{\mathbf{AVAR}}_m$	FGBL	FGBM	FESX	FDAX
04/23/2008	FGBL	1.55 (2,5)	0.64 (1,5)	1.68 (2,4)	1.22 (1,2)
	FGBM		0.27 (2,4)	0.95 (1,3)	0.35 (1,2)
	FESX			242.58 (2,5)	31.00 (1,2)
	FDAX				43.06 (2,5)
01/10/2008	FGBL	0.98 (2,12)	0.50 (1,2)	0.87 (2,3)	0.52 (1,2)
	FGBM		0.25 (2,11)	0.38 (1,3)	0.22 (1,2)
	FESX			86.34 (2,14)	7.06 (1,2)
	FDAX				24.43 (2,15)
09/11/2001	FGBL	18.22 (1,3)	2.21 (2,3)	311 (1,2)	121 (1,2)
	FGBM		5.71 (1,3)	173 (1,2)	668 (1,1)
	FESX			46700 (2,5)	27800 (1,1)
	FDAX				14800 (2,5)

Table 6.6: Estimates for the asymptotic variances ($\cdot\, 10^{10}$) and used bins and multiscale frequencies (K_N, M_N) for binwise estimators.

Date	Z (p-value)	FGBL	FGBM	FESX	FDAX
04/23/2008	FGBL	8.45 (0)	5.84 (0)	0.86 (0.390)	0.39 (0.697)
	FGBM		8.25 (0)	-0.06 (0.952)	0.79 (0.430)
	FESX			7.33 (0)	4.81 (0)
	FDAX				15.27 (0)
01/10/2008	FGBL	6.27 (0)	4.69 (0)	0.99 (0.322)	0.71 (0.478)
	FGBM		5.16 (0)	1.14 (0.254)	1.28 (0.201)
	FESX			9.85 (0)	12.28 (0)
	FDAX				17.47 (0)
09/11/2001	FGBL	6.88 (0)	7.57 (0)	-2.95 (0.003)	-3.10 (0.002)
	FGBM		7.25 (0)	-1.82 (0.069)	-0.89 (0.373)
	FESX			17.88 (0)	5.23 (0)
	FDAX				19.81 (0)

Table 6.7: Estimates $Z = \widehat{\langle X,Y\rangle}_T^{multi} / \sqrt{\widehat{\mathbf{AVAR}}_{multi}}$ and probabilities (p-values) that $|Z| \geq$ estimate under the null that $\mathbb{E}Z = 0$.

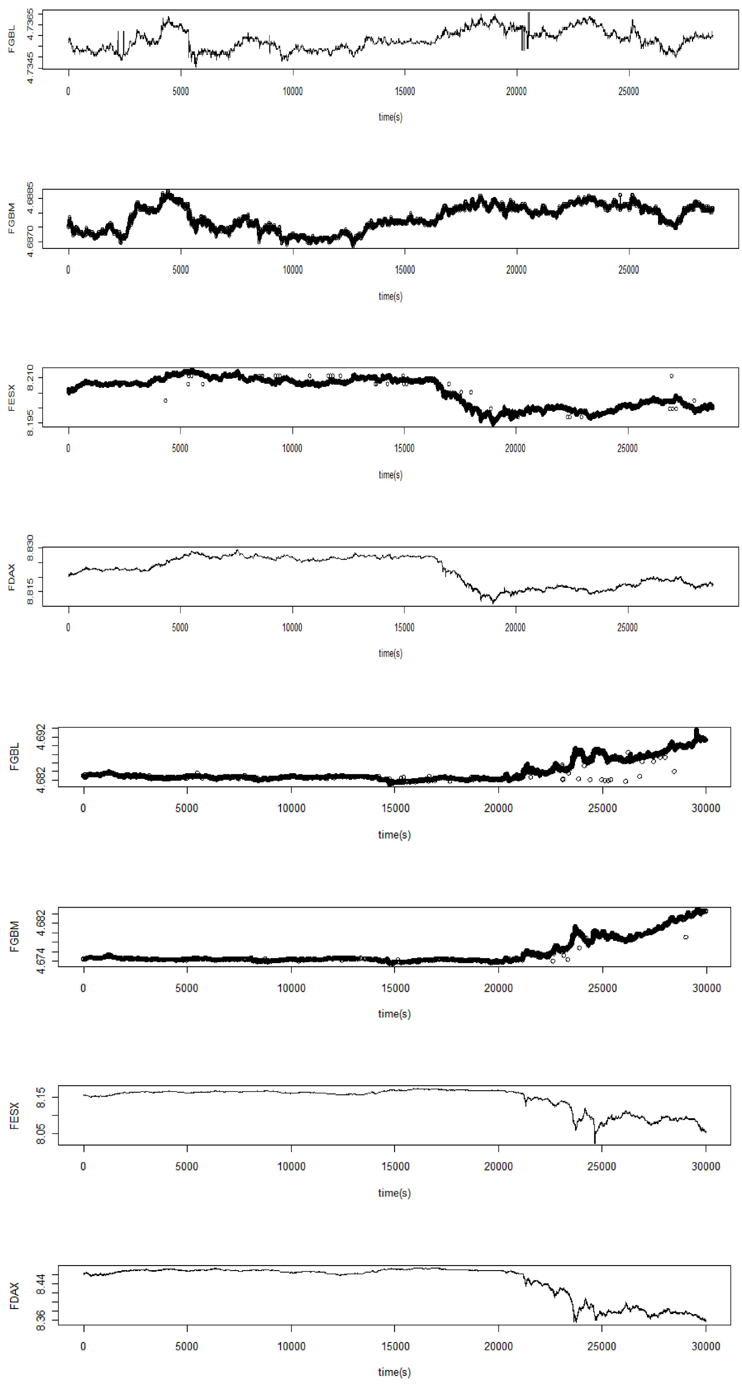


Figure 6.11: Sample paths of the four log-prices for 04/23/2008 (top) and 09/11/2001 (bottom).

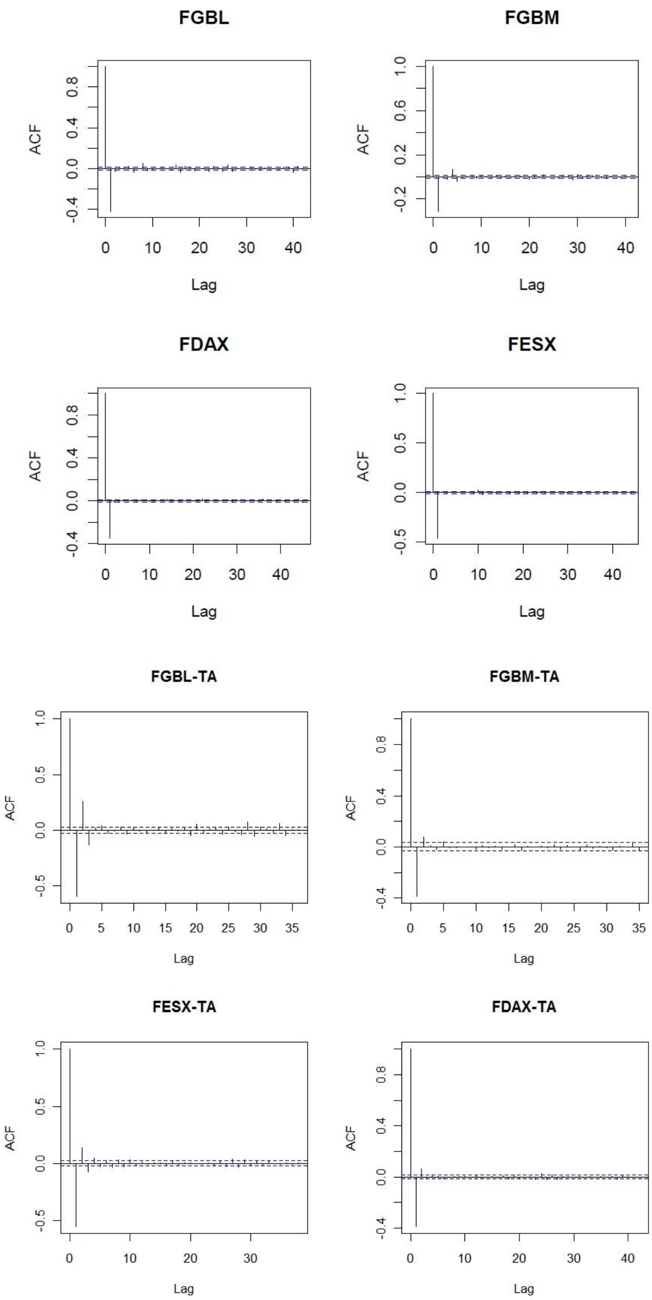


Figure 6.12: Autocorrelations of the tick (top) and the transaction-data (bottom) for 04/23/2008.

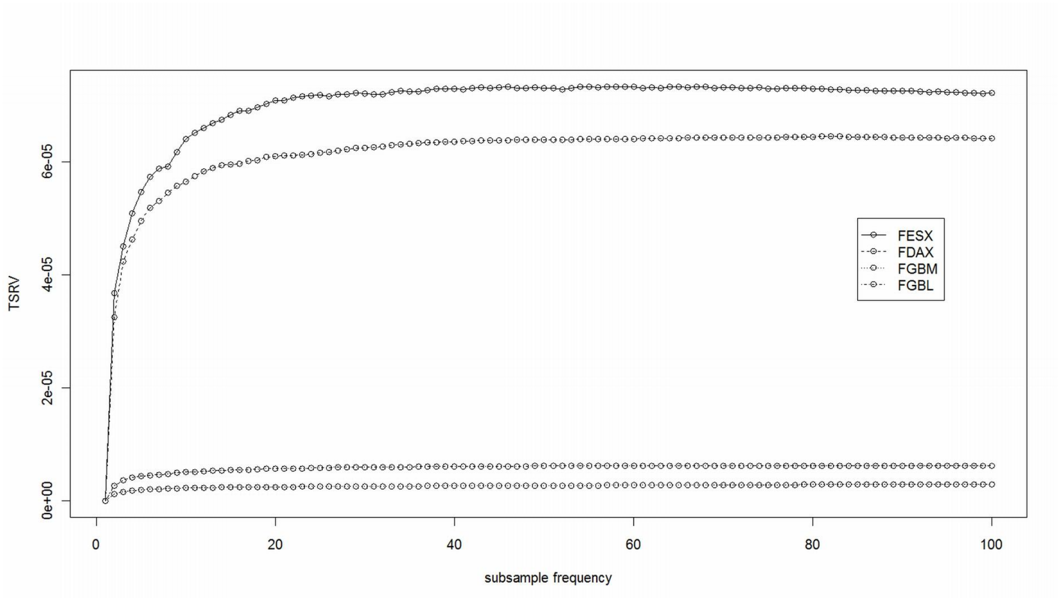


Figure 6.13: TSRV estimates for 04/23/2008.

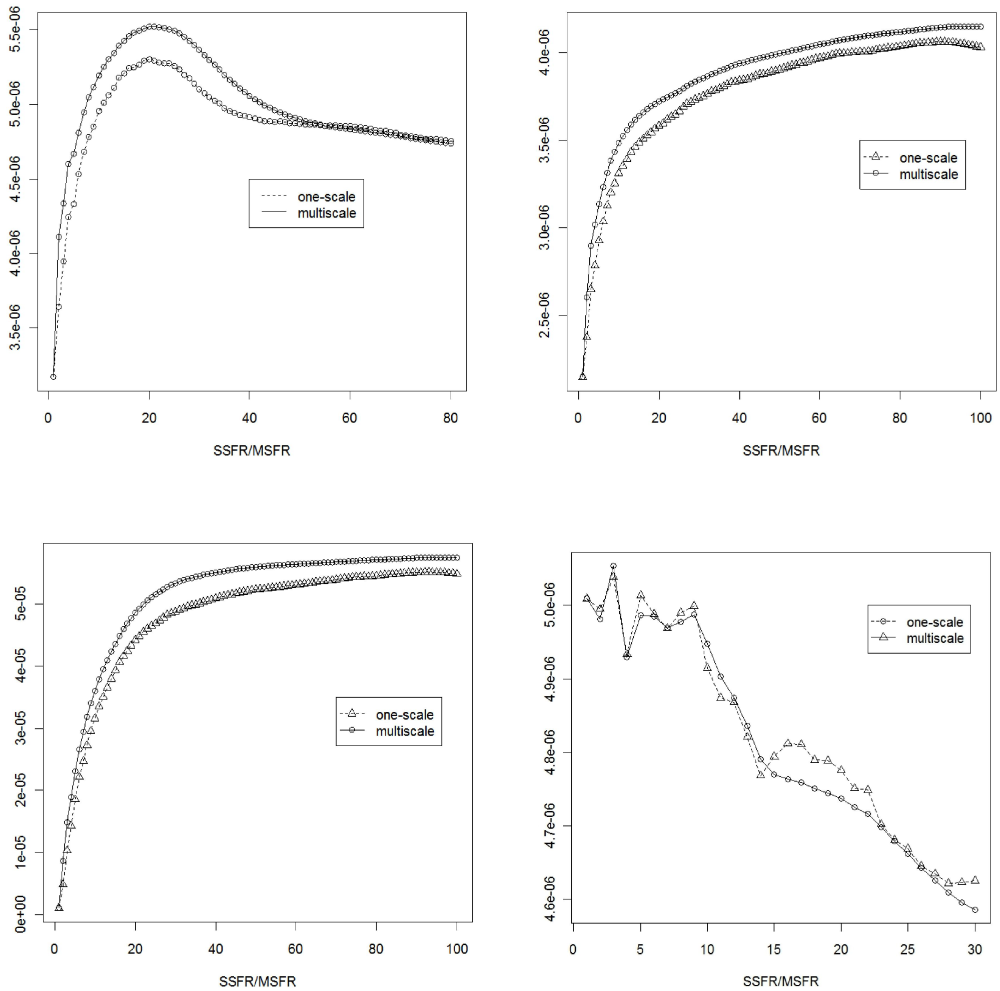


Figure 6.14: Generalized multiscale and one-scale subsampling estimators for ticks (left) and transaction data (right) FGBL/FGBM, 04/23/2008 (top), ticks FGBL/FGBM (left) and ticks FESX/FDAX (right), 01/10/2008 (bottom).

Bibliography

Yacine Aït-Sahalia, Jianqing Fan, and Dacheng Xiu. High-frequency estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association*, 105 (492):1504–1516, 2010.

Yacine Aït-Sahalia and Jean Jacod. Testing for jumps in a discretely observed process. *Annals of Statistics*, 37(1):184–222, 2009.

Yacine Aït-Sahalia, Lan Zhang, and Per A. Mykland. Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics,*, Forthcoming, 2009.

D.J. Aldous and G.K. Eagleson. On mixing and stability of limit theorems. *Annals of Probability*, 6(2):325–331, 1978.

Basel Awartani, Valentina Corradi, and Walter Distasor. Assessing market microstructure effects via realized volatility measures with an application to the dow jones industrial average stocks. *Journal of Business and Economic Statistics*, 27(2):251–265, 2009.

Ole E. Barndorff-Nielsen, Peter R. Hansen, Asger Lunde, and Neil Shephard. Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536, 2008a.

Ole E. Barndorff-Nielsen, Peter R. Hansen, Asger Lunde, and Neil Shephard. Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *SSRN working paper 1154144, University of Aarhus*, 2008b.

Ole E. Barndorff-Nielsen and Neil Shephard. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society*, 64(2):253–280, 2002.

Markus Bibinger. Efficient covariance estimation for asynchronous noisy high-frequency data. *Scandinavian Journal of Statistics*, 38:23–45, 2011.

Patrick Billingsley. *Probability and Measure*. Springer, New York, 2 edition, 1991.

Fischer Sheffey Black and Myron Samuel Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–654, 1973.

Richard C. Bradley. Basic properties of strong mixing conditions. a survey and some open questions. *Probability Surveys*, 2:107–144, 2005.

S.J. Brown. Estimating volatility. *in Figlewski, S., W. Silber and M. Subrahmanyam, Financial Options, : From Theory to Practice, Homewood, Illinois, Business One Irwin,*, pages 516–537, 1990.

K. C. Chan, Andrew Karolyi, Francis A. Longstaff, and Anthony B. Sanders. An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance*, 47:1209–1227, 1992.

Kim Christensen, Silja Kinnebrock, and Mark Podolskij. Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of Econometrics,*, Forthcoming, 2010.

David Roxbee Cox and Valerie Isham. *Point processes, Monographs on applied probability and statistics series.* Chapman and Hall, New York, 1980.

John C. Cox, Jonathan E. Ingersoll, and Stephen A. Ross. An analysis of variable rate loan contracts. *Journal of Finance*, 35:389–403, 1980.

Laurens de Haan and Ana Ferreira. *Extreme Value Theory: An Introduction.* Springer, New York, 2006.

Paul Doukhan. *Mixing: Properties and Examples, Lecture Notes in Statistics, vol.85.* Springer, Berlin, 1994.

Richard Durrett. *Probability: Theory and Examples.* Duxbury Press, 1995.

Thomas W. Epps. Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, 74:291–298, 1979.

Jianqing Fan and Yazehn Wang. Multi-scale jump and volatility analysis for high-frequency data. *Journal of the American Statistical Association*, 102(480):1349–1362, 2007.

Paul D. Feigin. Stable convergence of semimartingales. *Stochastic Processes and their Applications*, 19(1):125 – 134, 1985.

Arnauld Gloter and Jean Jacod. Diffusions with measurement errors 1 and 2. *ESAIM Probability and Statistics*, 5:225–242, 2001.

Jaroslav Hájek. A characterization of limiting distributions of regular estimates. *Probability Theory and Related Fields*, 14(4):323–330, 1970.

Jaroslav Hájek. Local asymptotic minimax and admissibility in estimation. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, pages 175–194, 1972.

P. Hall and C.C. Heyde. *Martingale Limit Theory and its Application.* Academic Press, Boston, 1980.

T. Hayashi and N. Yoshida. On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, 11(2):359–379, 2005.

T. Hayashi and N. Yoshida. Asymptotic normality of a covariance estimator for non-synchronously observed diffusion processes. *Annals of the Institute of Statistical Mathematics*, 60(2):367–406, 2008.

T. Hayashi and N. Yoshida. Nonsynchronous covariation process and limit theorems. *Stochastic Processes and their Applications*, In Press, Uncorrected Proof, 2011. ISSN 0304-4149. URL `http://www.sciencedirect.com/science/article/B6V1B-51S6XGF-1/2/1ae16a817d8155f698610f572fd79370`.

Steven L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6:327–343, 1993.

Xin Huang and George Tauchen. The relative contribution of jumps to total price variation. *Journal of Financial Econometrics*, 3:456–499, 2005.

Jean Jacod. On continuous conditional gaussian martingales and stable convergence in law. *Séminaire de Probabilitiés*, pages 232–246, 1997.

Jean Jacod, Yan Li, Per A. Mykland, Mark Podolskij, and Mathias Vetter. Microstructure noise in the continous case: the pre-averaging approach. *Stochastic Processes and their Applications*, 119:2803–2831, 2009.

Jean Jacod and Philip Protter. Asymptotic error distributions for the euler method for stochastic differential equations. *The Annals of Probability*, 26:267–307, 2003.

Jean Jacod and Albert N. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer, New York, 2003.

Ilze Kalnina and Oliver Linton. Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error. *Journal of Econometrics*, 147: 47–59, 2008.

Ioannis Karatzas and Stephen E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, New York, 2 edition, 1991.

Robert S. Lipster and Albert N. Shiryaev. *Statistics of Random Processes 1, General Theory*. Springer, Berlin, 2 edition, 2001.

Lucien Le Cam. Locally asymptotically normal families of distributions. *University of California Publications in Statistics*, pages 37–98, 1960.

Lucien Le Cam. Limits of experiments. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, 1:245–261, 1972.

Lucien Le Cam. *Asymptotic Methods in Statistcial Decision Theory.* Springer, New York, 1986.

Per Mykland and Lan Zhang. Inference for continuous semimartingales observed at high frequency. *Econometrica*, 77(5):1403–1445, 2009.

Bernt Øksendal. *Stochastic Differential Equations.* Springer, Berlin, 2005.

Alessandro Palandri. Consistent realized covariance for asynchronous observations contaminated by market microstructure noise. Technical report, University of Copenhagen, 2006.

Magda Peligrad. On the asymptotic normality of sequences of weakly dependent random variables. *Journal of Theoretical Probability*, 9(3):703–716, 1996.

Mark Podolskij and Mathias Vetter. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli*, 15(3):634–658, 2009.

Mark Podolskij and Mathias Vetter. Understanding limit theorems for semimartingales: a short survey. *Statistica Nederlandica*, 64(3):329–351, 2010.

Philip E. Protter. *Stochastic Integration and Differential Equations.* Springer, Berlin, 2004.

C. Radhakrishna Rao. *Linear Statistical Inference and its Applications.* Wiley, New York, 2 edition, 2001.

Markus Reiß. Asymptotic equivalence for inference on the volatility from noisy observations. *Annals of Statistics,*, Forthcoming, 2011.

Alfred Rényi. On stable sequences of events. *Sankhya: The Indian Journal of Statistics, Series A*, 25(3):293–302, 1963.

Daniel Revuz and Marc Yor. *Continous Martingales and Brownian Motion.* Springer, Berlin, 1991.

Stephen E. Shreve. *Stochastic Calculus for Finance II: Continuous-Time Models*. Springer, Berlin, 1 edition, 2008.

Sergey Utev. Central limit theorem for dependent random variables. *In Grigelionis et al., Probability theory and mathematical statistics*, 2:519–528, 1990.

Aad W. van der Vaart. *Asymptotic Statistics.* University Press, Cambridge, 1998.

Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes.* Springer, New York, 1996.

Harry van Zanten. A multivariate central limit theorem for continuous local martingales. *Statistics and Probability Letters*, 50:229 – 235, 2000.

Oldrich Alfons Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5:177–188, 1977.

Dacheng Xiu. Quasi-maximum likelihood estimation of volatility with high frequency data. *Journal of Econometrics*, 159:235–250, 2010.

Lan Zhang. Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli*, 12(6):1019–1043, 2006.

Lan Zhang, Per A. Mykland, and Yacine Aït-Sahalia. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100(472):1394–1411, 2005.

List of Figures

List of Tables

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den 30.03.2011 Markus Bibinger